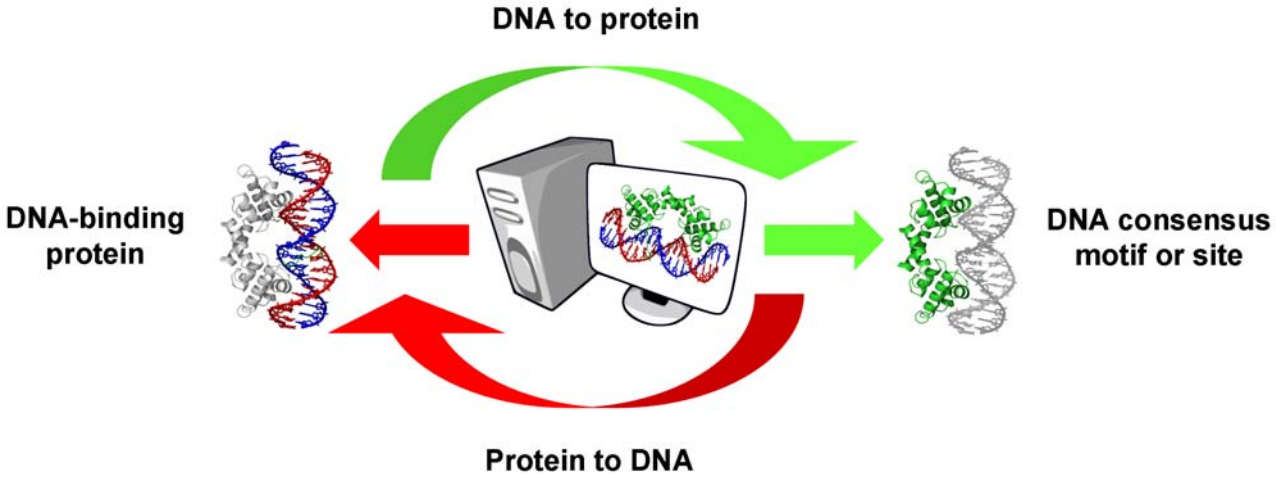
footprintDB

http://floresta.eead.csic.es/footprintdb



User Manual

Alvaro Sebastian Yagüe & Bruno Contreras-Moreira

Laboratory of Computational Biology

Estación Experimental de Aula Dei / CSIC

Av. Montañana 1.005, 50059 Zaragoza (SPAIN)

Index

Index ____ *2*

Introduction ____ *3*

1. **footprintDB is a repository of databases** ____ **3**
2. **Annotation of transcription factor interfaces** ____ **4**
3. **footprintDB is a search engine** ____ **5**

Web site navigation ____ *6*

1. **Sections** ____ **6**
2. **Navigation** ____ **7**

User registration ____ *8*

3. **Registration** ____ **8**
4. **Log In** ____ **8**
5. **Log out** ____ **9**
6. **Recover account info** ____ **9**
7. **Modify account info** ____ **10**
8. **Delete account** ____ **10**

Searching ____ *11*

1. **Search keywords** ____ **11**
2. **Search DNA motifs** ____ **14**
3. **Search protein sequences** ____ **24**
4. **Retrieve stored searches** ____ **28**
5. **Searching through the Web services interface** ____ **29**

Database insertion ____ *33*

1. **Insert a new database into footprintDB** ____ **33**
2. **footprintDB and TRANSFAC data formats** ____ **34**
3. **Manage own databases** ____ **37**

Introduction

footprintDB is a **web server for assigning putative *cis* DNA motifs to input transcription factors (TFs) and conversely for predicting which TFs that might recognize input DNA motifs**.

footprintDB predictions can be extended to external proteomes to design DNA binding experiments for the desired organism.

footprintDB database consists of a collection of **curated and annotated DNA binding data**, which is obtained **from literature and public repositories** and stored in a database. Among these data are the protein sequences of the **TFs**, their **DNA binding sites (DBSs)** and their **Position-Specific Scoring Matrices (PSSM)** that summarize the binding preferences, together with their Pfam protein domains, literature references and the set of annotated DNA binding protein **interface residues**.

footprintDB features are described in more detail in the following sections.

1. footprintDB is a repository of databases

Current online release of **footprintDB contains 2422 unique TF sequences, 3662 PSSMs and 10112 DBSs**.

footprintDB is by design a meta-database of TFs attached to their experimentally determined DNA binding preferences (PSSMs and DBSs). Therefore it does not incorporate other databases which contain only TF, DBS or predicted regulatory sequences. **The first building block is 3D-footprint** (Contreras-Moreira 2010), a database for the structural analysis of protein-DNA complexes, for two reasons: i) it is to our knowledge the only up-to-date source of annotated binding interfaces of TFs; and ii) it contains structure-based PSSMs, motifs inferred from *cis* elements captured in X-ray and NMR complexes, that have been independently validated (AlQuraishi and McAdams 2011; Lin and Chen 2013). The remaining databases and repositories integrated in footprintDB are:

(i) **JASPAR CORE** (2009 version, all species redundant set): a high-quality collection of transcription factor DNA-binding preferences, modeled as PSSMs (Portales-Casamar, Thongjuea et al. 2010).

(ii) **UniPROBE** (Universal PBM Resource for Oligonucleotide Binding Evaluation, Sep 2012 version): contains *in vitro* DNA binding specificities of proteins measured with universal protein binding microarrays (Robasky and Bulyk 2011).

(iii) **"HumanTF"**: sequence-specific binding preferences of human TFs obtained by high-throughput SELEX and ChIP sequencing. It includes a total of 830 binding profiles, describing 239 distinctly different binding specificities (Jolma, Yan et al. 2013).

(iv) **Athamap**: genome-wide map of potential transcription factor binding sites (TFBS) in Arabidopsis thaliana (Bulow, Engelmann et al. 2009).

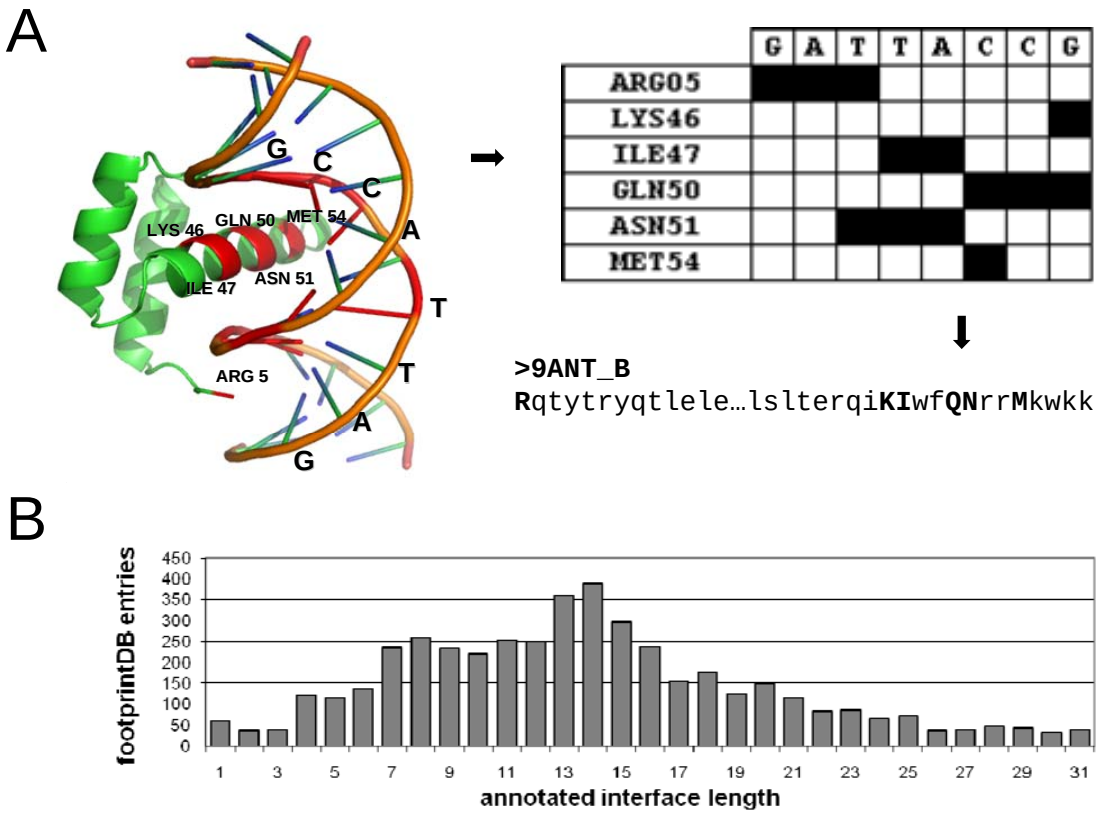
(v) **RegulonDB** (7.5 version): contains curated data of the transcriptional regulatory network of Escherichia coli K12, including PSSMs and DBSs for many TFs (Salgado, Peralta-Gil et al. 2013).

(vi) **DBTBS** (Database of transcriptional regulation in *Bacillus subtilis*): A database of transcriptional regulation in *Bacillus subtilis* (Sierro, Makita et al. 2008).

(vii) **"DrosophilaTF"**: Motifs for 56 *Drosophila melanogaster* transcription factors built from *in vitro* binding site selection experiments and compiled genomic binding site sequences (Down, Bergman et al. 2007).

2. Annotation of transcription factor interfaces

TF sequences in footprintDB have their DNA-binding interfaces annotated by means of BLASTP alignments against the 3D-footprint library (http://floresta.eead.csic.es/3dfootprint/download/list_interface2dna.txt) with an E-value threshold of 10. Aligned interface positions from one or more protein-DNA complexes are thus transferred to entries in the database like explained in the following Figure.



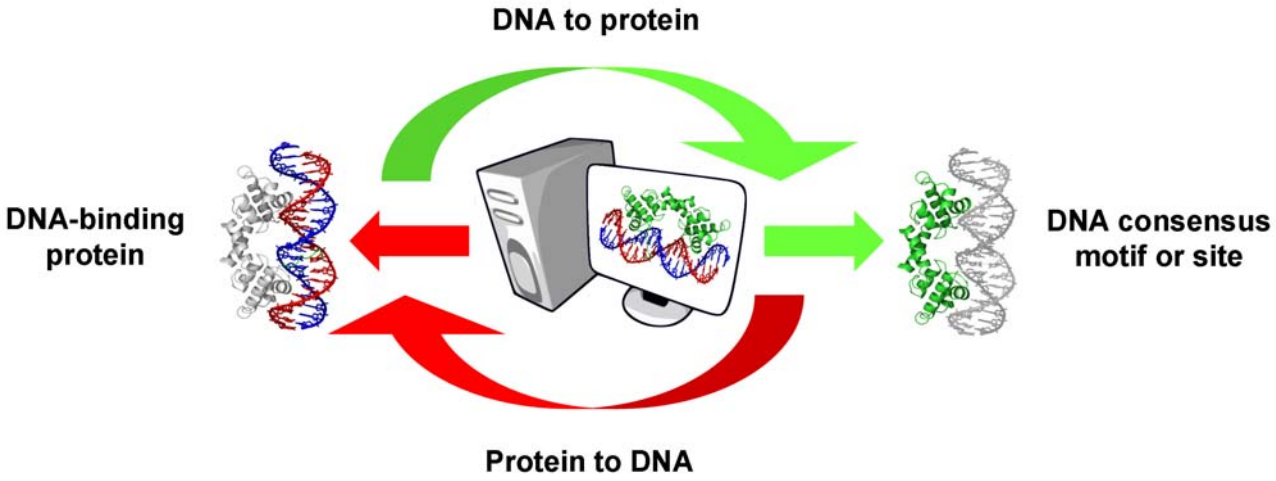
Annotation of interface residues applying the geometrical rules of 3D-footprint. **(A)** Interface of PDB entry 9ANT, which corresponds to Homebox protein *Antennapedia* in complex with a *cis* element. First, inter-atomic distances are calculated among heavy atoms of both amino acid side chains and nitrogen bases. Second, a matrix of interface contacts is compiled. Third, interface residues are marked as upper-case letters in the sequence. **(B)** Histogram of predicted interfaces in footprintDB, transferred from 3D-footprint entries through BLASTP alignments.

3. footprintDB is a search engine

The footprintDB search engine is designed primarily to receive two types of queries:

1. INPUT: a DNA consensus motif or site
   OUTPUT: a list of DNA-binding proteins (mainly TFs) predicted to bind a similar DNA motif

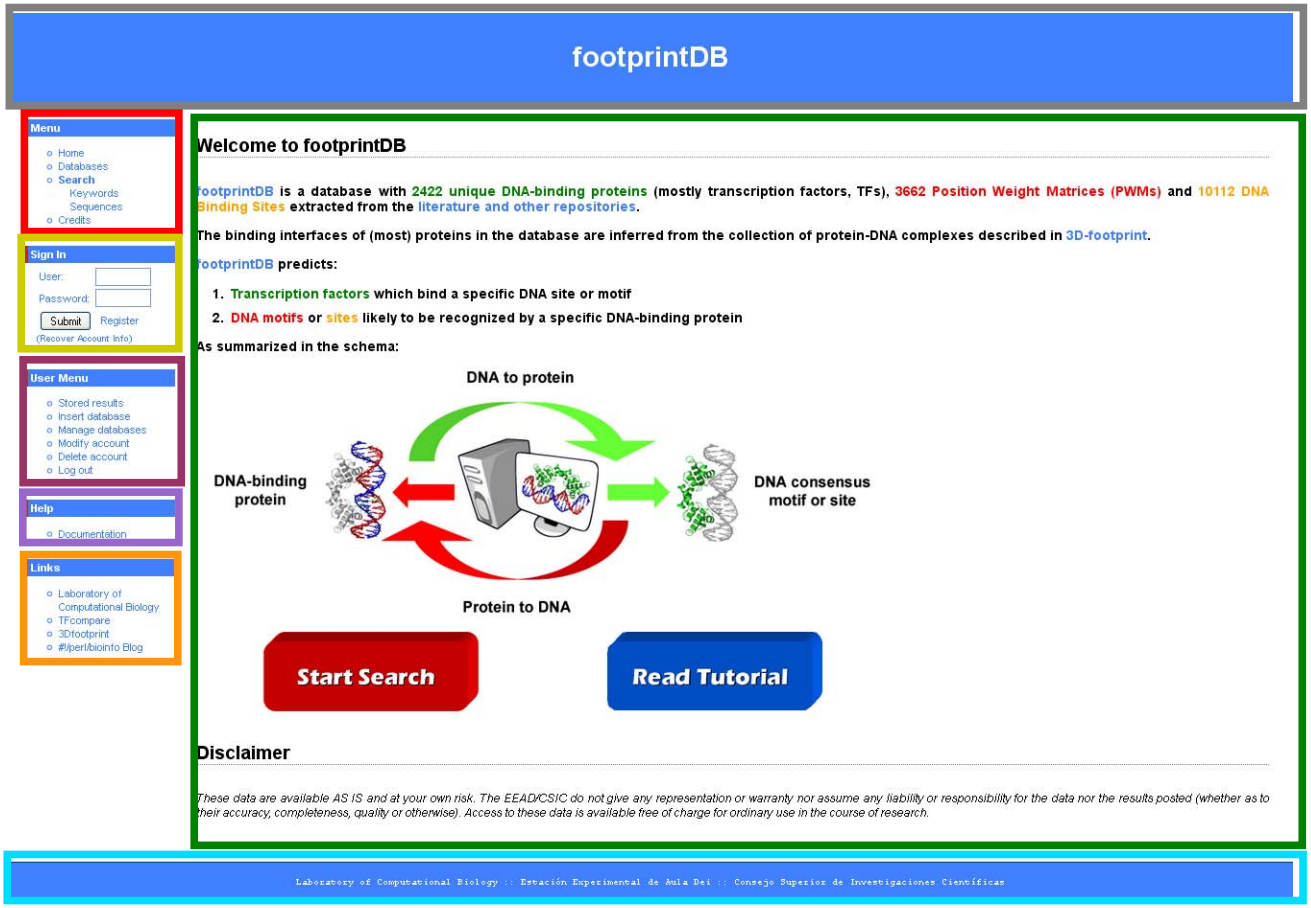
2. INPUT: a protein sequence of a putative DNA-binding protein
   OUTPUT: a list of possibly recognized DNA motifs



Flowchart of the footprintDB search engine.

Web site navigation

1. Sections


footprintDB
Menu
Home
Databases
Search
Keywords
Sequences
Credits
Sign In
User:
Password:
Submit
Register
(Recover Account Info)
User Menu
Stored results
Insert database
Manage databases
Modify account
Delete account
Log out
Help
Documentation
Links
Laboratory of Computational Biology
TFcompare
3Dfootprint
#/perl/bioinfo Blog
Welcome to footprintDB
footprintDB is a database with 2422 unique DNA-binding proteins (mostly transcription factors, TFs), 3662 Position Weight Matrices (PWMs) and 10112 DNA Binding Sites extracted from the literature and other repositories.
The binding interfaces of (most) proteins in the database are inferred from the collection of protein-DNA complexes described in 3D-footprint.
footprintDB predicts:
1. Transcription factors which bind a specific DNA site or motif
2. DNA motifs or sites likely to be recognized by a specific DNA-binding protein
As summarized in the schema:
DNA to protein
DNA-binding protein
DNA consensus motif or site
Protein to DNA
Start Search
Read Tutorial
Disclaimer
These data are available AS IS and at your own risk. The EEAD/CSIC do not give any representation or warranty nor assume any liability or responsibility for the data nor the results posted (whether as to their accuracy, completeness, quality or otherwise). Access to these data is available free of charge for ordinary use in the course of research.
Laboratory of Computational Biology :: Estación Experimental de Aula Dei :: Consejo Superior de Investigaciones Científicas

From top to bottom and from left to right:

Main menu: links to Home, Database listing, Search Keywords, Search sequences and Credits sections.

Sign In menu: Authentication form and Registration form link.

User menu: links to Stored Results, Insert Databases, Manage Databases, Modify Account, Delete Account and Log out.

Help menu: link to Documentation Section.

Links menu: links to recommended Internet resources.

2. Navigation

The menus on the left side can be used to navigate across the site.

'Main menu' is composed by the following sections:

- **Home**: Access to home page with general information.
- **Databases**: Updated information about the databases included in footprintDB.
- **Search**: Access to search forms.
    - **Keywords**: to search keywords and data accessions or identifiers
    - **Sequences**: to search DNA motifs and TF protein sequences
- **Credits**: Information about footprintDB creators, citing, data sources and other resources.

'Sign In menu' is composed by an authentication form and a couple of links to:

- **Register**: Access to a registration form for new users.
- **Recover Account Info:** registered users can recover their account data..

'User menu' is only visible for authenticated users and is composed by the following sections:

- **Stored results**: Access to a historical record of searches performed by the user.
- **Insert database**: Insertion of user data collections.
- **Manage databases**: Manage user data inserted previously.
- **Modify account**: A form to modify user account data.
- **Delete account**: An option to remove an account.
- **Log out**

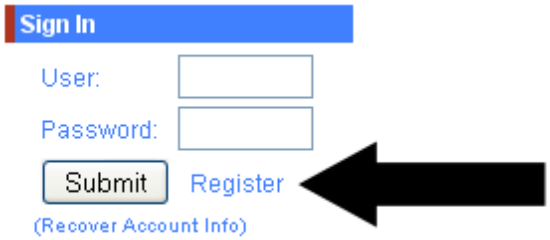
'Help menu' provides links to extensive footprintDB documentation.

'Links menu' provides links to our Laboratory of Computational Biology and other related links.

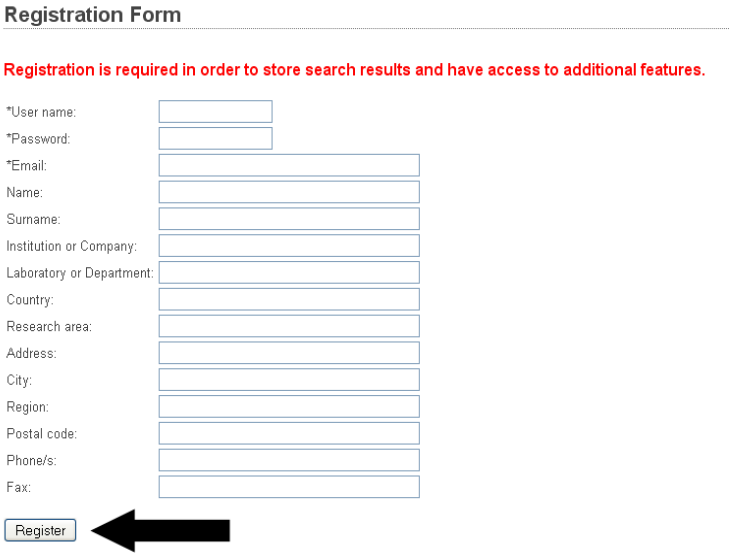
User registration

3. Registration

Click on the link 'Register' in the 'Sign In' menu on the left or go directly to http://floresta.eead.csic.es/footprintdb/index.php?user_register



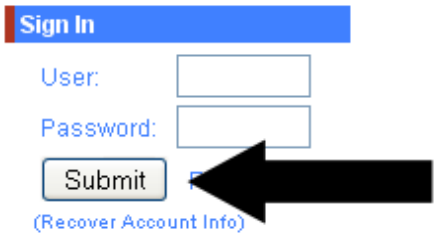
Fill in the registration form, required fields are marked with asterisk, and push the 'Register' button to submit the data:



You will see the following message if registration was successful: **"User successfully registered, you will shortly receive an email with your account information"** . An email is sent to remember your account data.

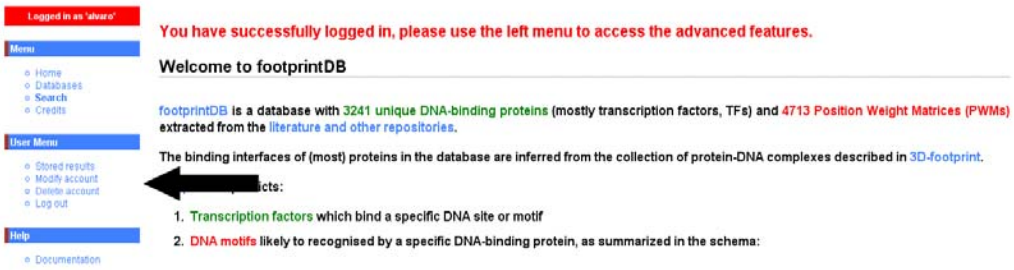
4. Log In

Enter 'User' and 'Password' in the 'Sign In' menu on the left and push the 'Submit' button:



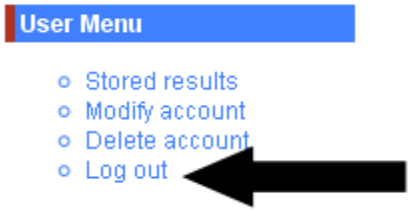
If successful, a message will be shown, your user name will be shown in red in the top of

the left menus and a the 'User Menu' will appear:


Logged in as 'alvaro'
Menu
Home
Databases
Search
Credits
User Menu
Stored results
Modify account
Delete account
Log out
Help
Documentation
You have successfully logged in, please use the left menu to access the advanced features.
Welcome to footprintDB
footprintDB is a database with 3241 unique DNA-binding proteins (mostly transcription factors, TFs) and 4713 Position Weight Matrices (PWMs) extracted from the literature and other repositories.
The binding interfaces of (most) proteins in the database are inferred from the collection of protein-DNA complexes described in 3D-footprint.
1. Transcription factors which bind a specific DNA site or motif
2. DNA motifs likely to recognised by a specific DNA-binding protein, as summarized in the schema:

5. Log out

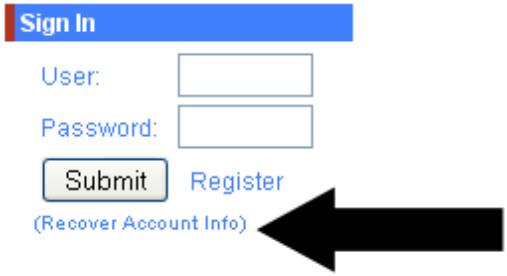
Click on the 'Log out' option in the 'User menu' on the left side:



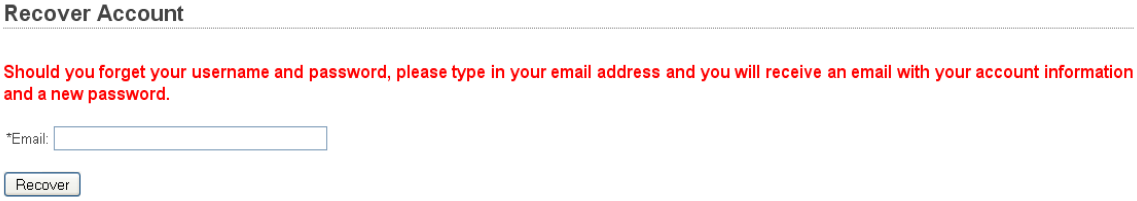
You will see the following message: **"You have successfully logged out, thank you or using footprintDB"** and 'User Menu' will hide (unless automatic cache is activated in your browser; in this case the menu will be visible until any other item is clicked).

6. Recover account info

Click on the 'Recover account info' link in the 'Log In' menu on the left side:

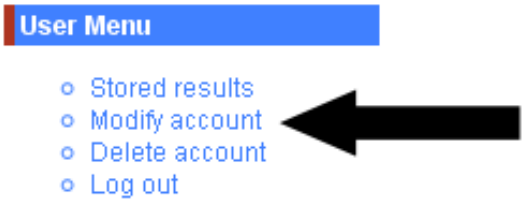


Enter your email address and push the 'Recover' button. If any account is associated to that address, you will receive your account data by email and a new auto-generated password.


Recover Account
Should you forget your username and password, please type in your email address and you will receive an email with your account information and a new password.
*Email:
Recover

7. Modify account info

Log in and click on the 'Modify account' link in the 'User menu' on the left side:

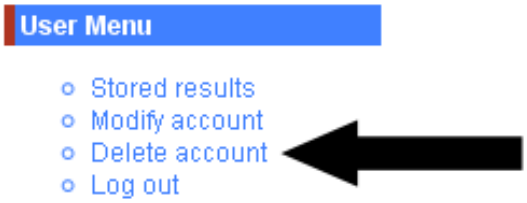


Modify your data in the formulary, required fields are marked with asterisk, and push the 'Modify' button to submit the data.

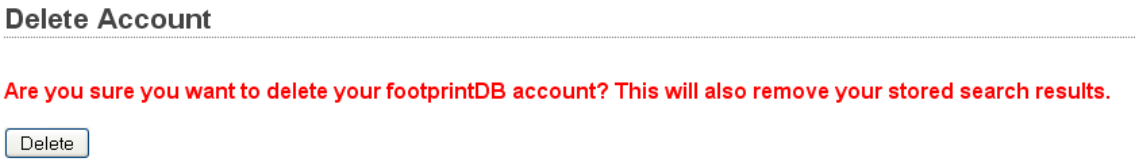
You will see the following message if registration was successful: **"User account successfully modified, you will shortly receive an email with your new account information."** and you will receive an email to remember your account data.

8. Delete account

Log in and click on the 'Delete account' link in the 'User menu' on the left side:



Please confirm that you want to delete the account by pushing the 'Delete' button.


Delete Account
Are you sure you want to delete your footprintDB account? This will also remove your stored search results.
Delete

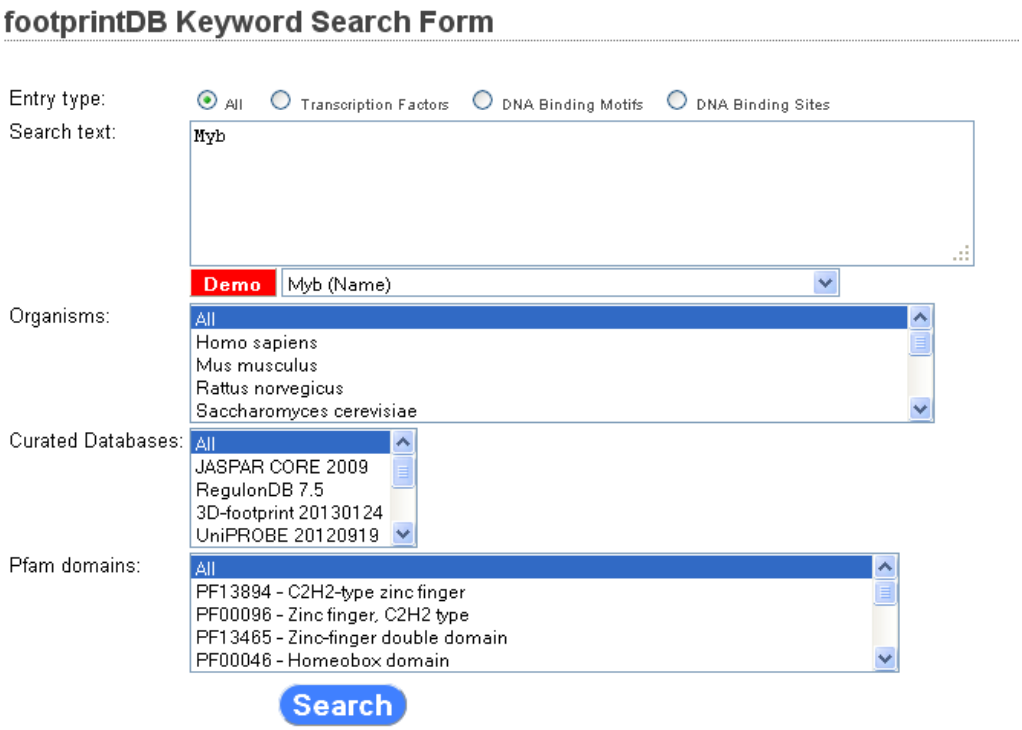
Searching

1. Search keywords

If you have a footprintDB account, log in first into the website to store your searches and reuse them; if you haven't got one, registration is recommended.

Click in the 'Search Keywords' option of the 'Main menu' or go directly to the url: http://floresta.eead.csic.es/footprintdb/index.php?search_entries

The search form looks like this:



The search form has the following fields and options:

Entry type: To restrict search to 'Transcription Factors', 'DNA Binding Motifs' or 'DNA Binding Sites'

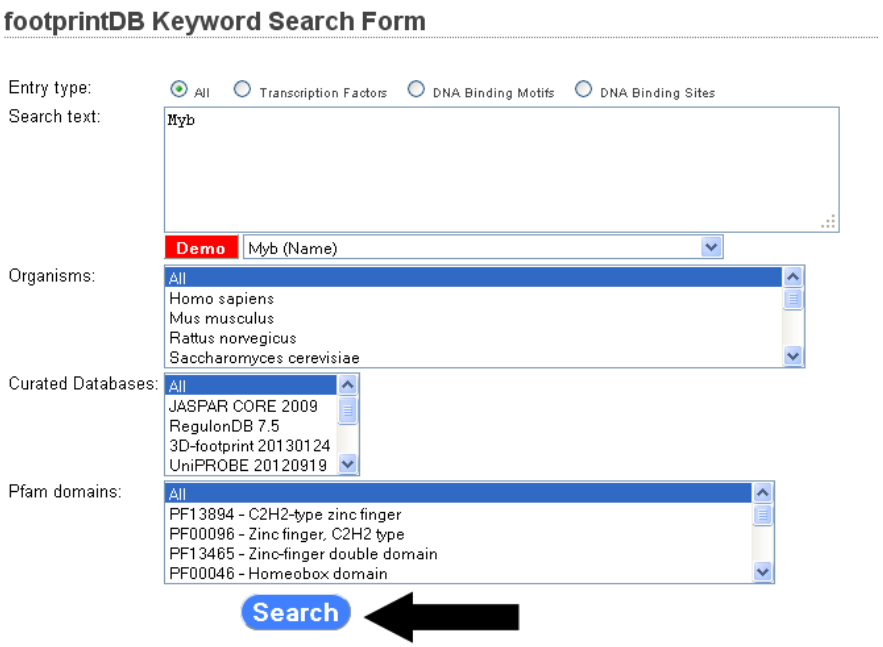
Search text: Text to search, it can be any descriptive word, a transcription factor protein or gene name, UniProt or PDB identifier, original source accession name or DNA site sequence.

Organisms: Select any organism(s) to restrict the search. Multiple species can be selected pushing the Control key on your keyboard. (Use with caution, as many TFs are not associated to an specific organism)

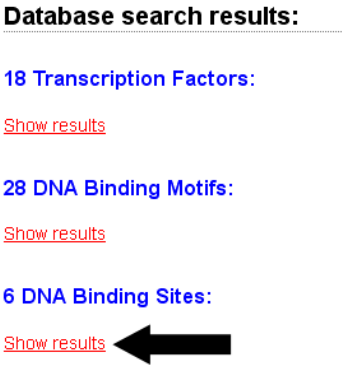
Databases: Select databases or sources to restrict the search. Multiple databases can be selected pushing the Control key.

Pfam domains: Select protein Pfam domains to restrict the search. Multiple domains can be selected pushing the Control key.

To start the search please click the 'Demo' button and then the 'Search' button:



The former search will look for the word 'Myb' in the database, obtaining multiple results that we can expand clicking on 'Show results':



A full list of the results will be shown, with a short summary of them and the option to access them individually or download them:

Database search results:

18 Transcription Factors:

Show results

28 DNA Binding Motifs:

Show results

6 DNA Binding Sites:

Show results

Accessions	Sequence	Organisms	Binding Motifs	Binding TFs
MYB1_1 (Athamap 20091028) MYB2_3 (Athamap 20091028)	acaccctAACtgacacacattct	Arabidopsis thaliana	MYB1 yCTAACTG MYB2 vhAACbvm	MYB1 (Myb-like DNA-binding domain, Myb-like DNA-binding domain) MYB2 (Myb-like DNA-binding domain, Myb-like DNA-binding domain)
MYB1_2 (Athamap 20091028) MYB2_4 (Athamap 20091028)	tttcggcacgtctAACtgcgactggcag	Arabidopsis thaliana	MYB1 yCTAACTG MYB2 vhAACbvm	MYB1 (Myb-like DNA-binding domain, Myb-like DNA-binding domain) MYB2 (Myb-like DNA-binding domain, Myb-like DNA-binding domain)
MYB2_1 (Athamap 20091028)	gggaaAACcaaatccg	Arabidopsis thaliana	MYB2 vhAACbvm	MYB2 (Myb-like DNA-binding domain, Myb-like DNA-binding domain)
MYB2_2 (Athamap 20091028)	ggctagcAACgccaagtgg	Arabidopsis thaliana	MYB2 vhAACbvm	MYB2 (Myb-like DNA-binding domain, Myb-like DNA-binding domain)
MYB44_1 (Athamap 20091028)	agcttaatgtgtgtcaGTTAgggtgtctcg	Arabidopsis thaliana	MYB44 GTyAGTTAsG	MYB44 (Myb-like DNA-binding domain, Myb-like DNA-binding domain)
MYB44_2 (Athamap 20091028)	agcttcaaaaGTTAGTTAcg	Arabidopsis thaliana	MYB44 GTyAGTTAsG	MYB44 (Myb-like DNA-binding domain, Myb-like DNA-binding domain)

Download

If we click in the Accession of any of the results we are shown the individual data for it:

DNA Binding Site

Accessions: MYB1_1 (Athamap 20091028), MYB2_3 (Athamap 20091028)

Organisms: Arabidopsis thaliana

Libraries: Athamap 20091028 [1]

[1] Bülow L, Engelmann S, Schindler M, Hehl R. AthaMap, integrating transcriptional and post-transcriptional data. Nucleic acids research 37:D983-6 (2009). [Pubmed]

Length: 23

Sequence: acaccctAACtgacacacattct

Binding TFs: MYB1 (Myb-like DNA-binding domain, Myb-like DNA-binding domain)
MYB2 (Myb-like DNA-binding domain, Myb-like DNA-binding domain)

Binding Motifs: MYB1 yCTAACTG
MYB2 vhAACbvm

Publications: Urao T., Yamaguchi-Shinozaki K., Urao S., Shinozaki K. An Arabidopsis myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. Plant Cell 5:1529-1539 (1993). [Pubmed]

Hoeren FU, Dolferus R, Wu Y, Peacock WJ, Dennis ES. 1998. Evidence for a role for AtMYB2 in the induction of the Arabidopsis alcohol dehydrogenase gene (ADH1) by low oxygen. Genetics. 149:479-90. [Pubmed]

Download

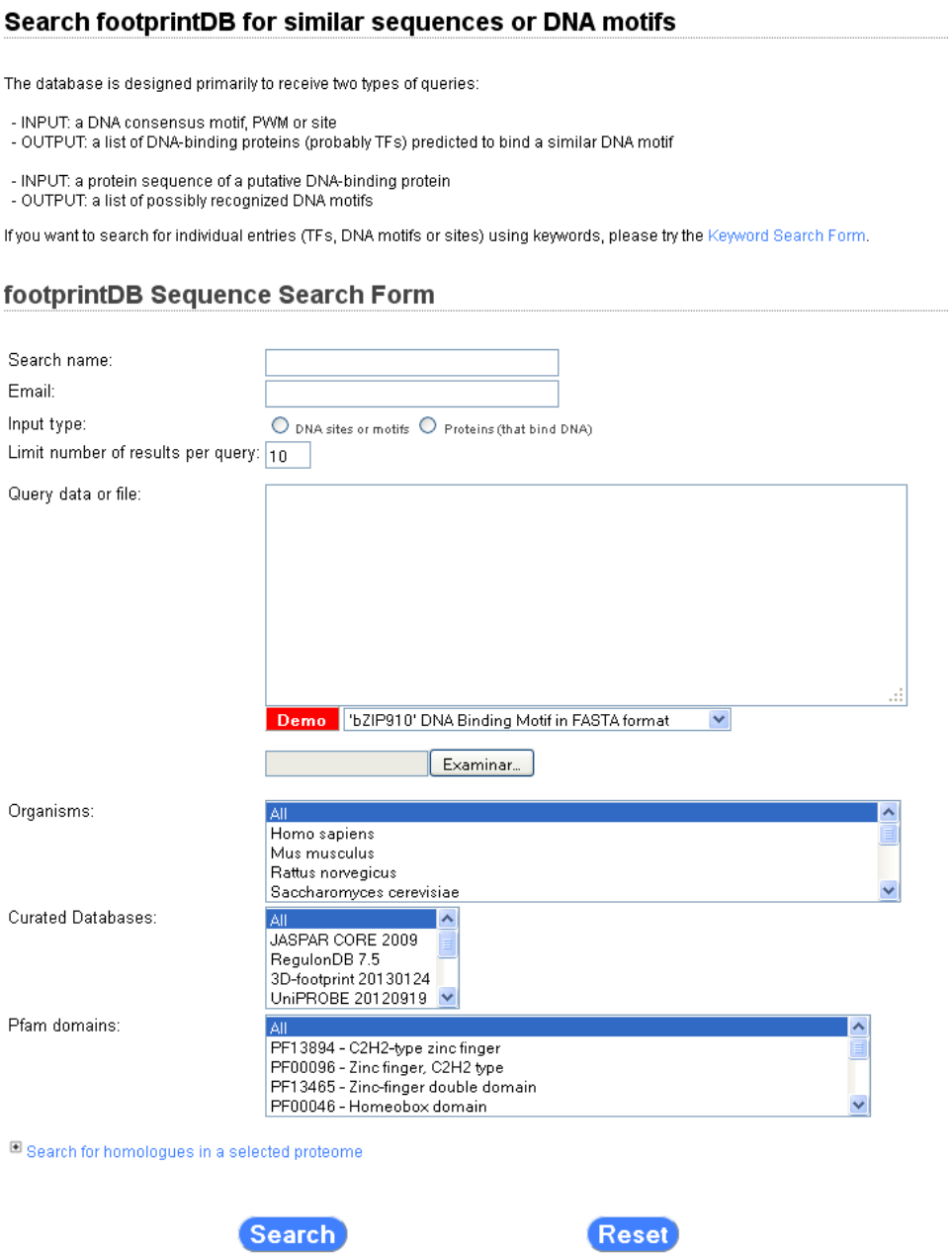
2. Search DNA motifs

a. Find transcription factors that bind DNA motifs similar to the query

If you have a footprintDB account, log in first into the website to store your searches and reuse them; if you haven't got one, registration is recommended.

Click on the 'Start to Search' button in the Home page, click in the 'Search sequences' option of the 'Main menu' or go directly to the url:
http://floresta.eead.csic.es/footprintdb/index.php?search

The search form looks like this:


Search footprintDB for similar sequences or DNA motifs
The database is designed primarily to receive two types of queries:
- INPUT: a DNA consensus motif, PWM or site
- OUTPUT: a list of DNA-binding proteins (probably TFs) predicted to bind a similar DNA motif
- INPUT: a protein sequence of a putative DNA-binding protein
- OUTPUT: a list of possibly recognized DNA motifs
If you want to search for individual entries (TFs, DNA motifs or sites) using keywords, please try the Keyword Search Form.
footprintDB Sequence Search Form
Search name:
Email:
Input type:
DNA sites or motifs
Proteins (that bind DNA)
Limit number of results per query:
10
Query data or file:
Demo
'bZIP910' DNA Binding Motif in FASTA format
Examinar...
Organisms:
All
Homo sapiens
Mus musculus
Rattus norvegicus
Saccharomyces cerevisiae
Curated Databases:
All
JASPAR CORE 2009
RegulonDB 7.5
3D-footprint 20130124
UniPROBE 20120919
Pfam domains:
All
PF13894 - C2H2-type zinc finger
PF00096 - Zinc finger, C2H2 type
PF13465 - Zinc-finger double domain
PF00046 - Homeobox domain
Search for homologues in a selected proteome
Search
Reset

The search form has the following fields and options:

Search name: Name a title for the search.

Email: Please type your email if you desire to receive the results by email.

Input type: Please choose **‘DNA sites or motifs’**..

Limit number of results per query: Enter the number of desired results per query.

Order results by: Allows to order results by DNA or TF similarity or by E-value.

Color results using twilight thresholds: Only available for DNA search, mark in green color results that pass the thresholds defined in our previous article and in red if not (Sebastian and Contreras-Moreira 2013).

Query data or file: Enter your DNA sites or motifs in the text area or upload them from a

file. The only valid formats are: FASTA and TRANSFAC. You can also use sample data pushing the ‘Demo’ button.

Table 1. Examples of FASTA and TRANSFAC formats for DNA input.

DNA motif in FASTA format:	DNA motif in TRANSFAC format:
>bZIP910 (JASPAR CORE) ATGACGT CTGACGT ATGACGT CTGACGT GTGACGT GTGACGT …	DE bZIP910 (JASPAR CORE) 1 15 15 5 0 2 0 0 0 35 3 0 0 35 0 4 35 0 0 0 5 0 35 0 0 6 0 0 35 0 7 0 0 0 35 XX

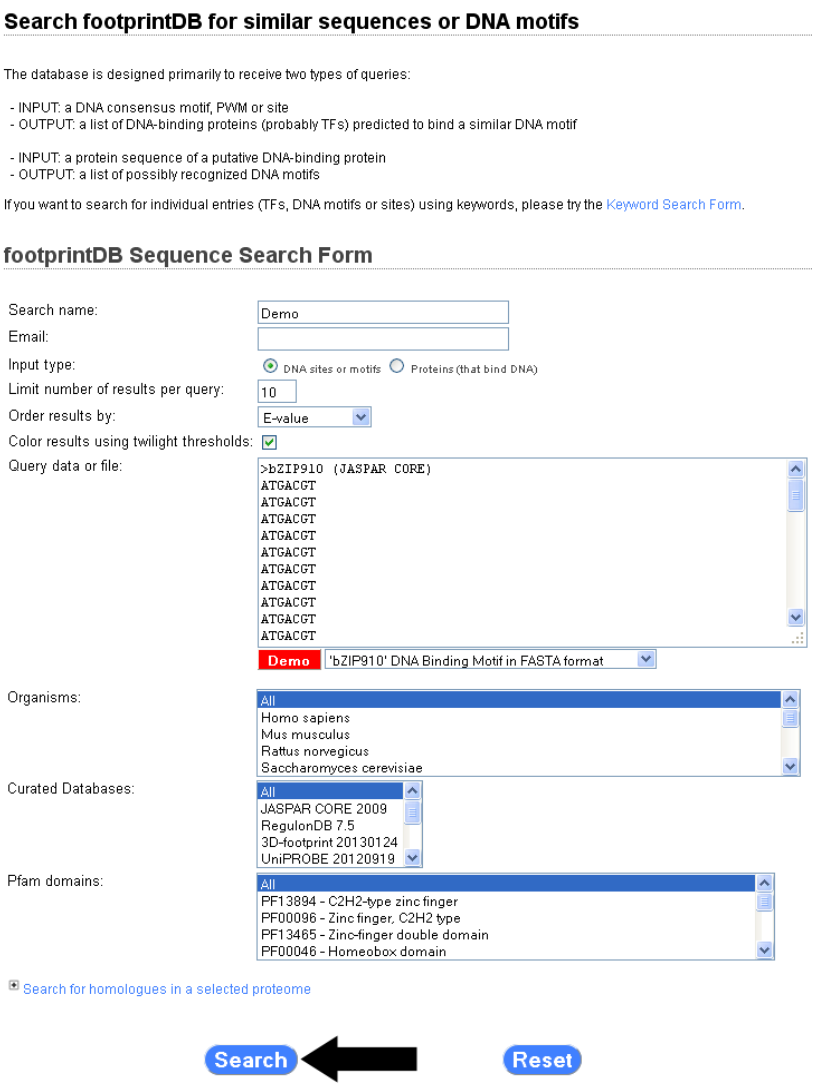
Organisms: Select any organism(s) to restrict the search. Multiple species can be selected pushing the Control key on your keyboard. (Use with caution, as many TFs are not associated to an specific organism)

Databases: Select databases or sources to restrict the search. Multiple databases can be selected pushing the Control key.

Pfam domains: Select protein Pfam domains to restrict the search. Multiple domains can be selected pushing the Control key.

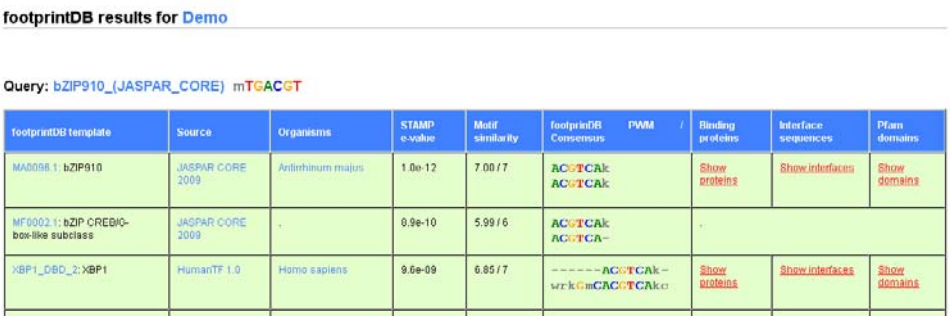
Please not that when search type **‘DNA sites or motifs’** is selected, the option to automatically ‘Search **for homologues in a selected proteome’** is shown, which will be explained in the next section

To start the search please click on the **‘Demo’** button and then on **‘Search’** :



In the former search we take as query a DNA motif in FASTA format (a list of DNA binding sites, all of them with the same length). We want to search at most 10 TFs with similar DNA motifs without filtering organisms neither domains in all available databases.

We obtain the following results (only the first 3 are shown):

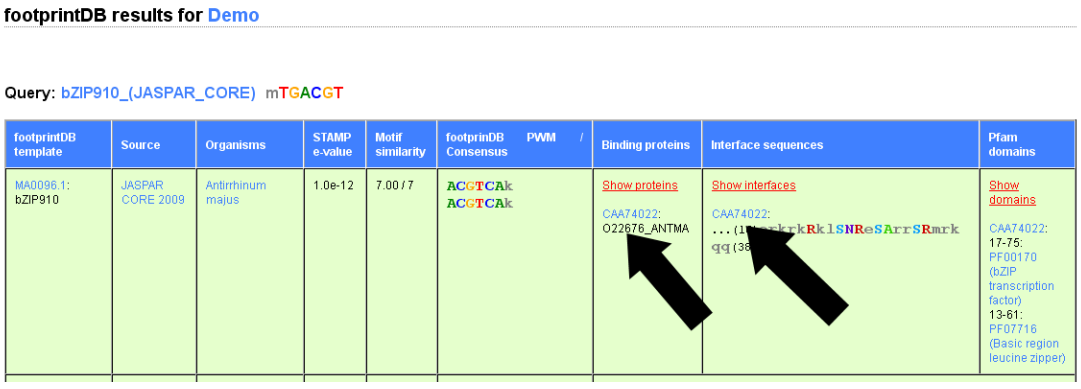


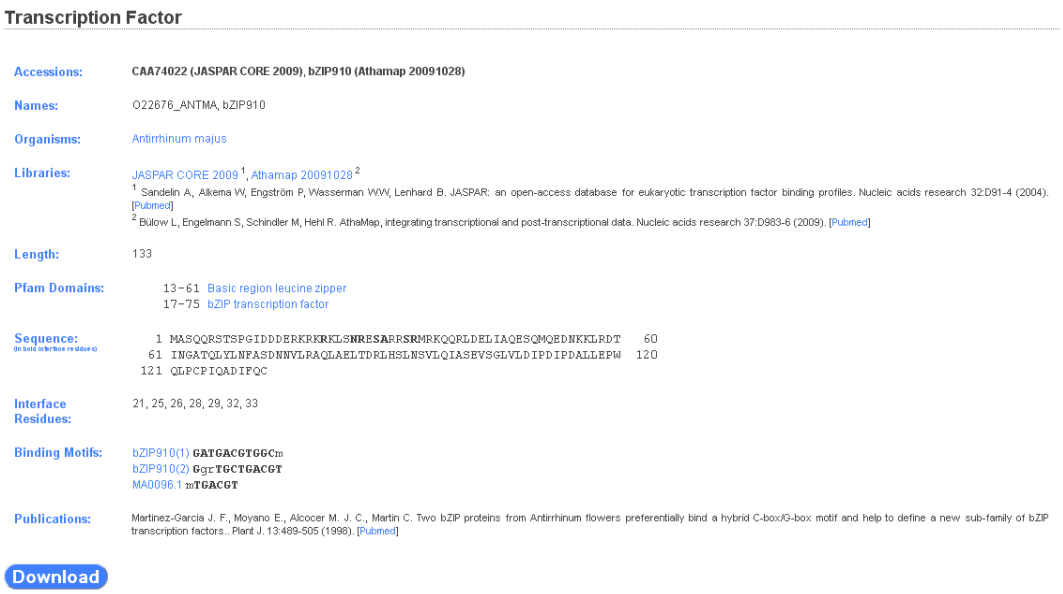
We notice that the first result is the query itself (because Demo query is from JASPAR database included in footprintDB) and the others are similar DNA motifs present in footprintDB.

We can click on the links **'Show proteins', 'Show interfaces' and 'Show domains'** to retrieve information about proteins that bind the similar DNA domain retrieved in the search (when there are annotated TFs for the DNA motifs, second result has not related TF).

Predicted DNA binding residues for each protein are shown coloured in the interface sequence. Left-clicking on the 'footprintDB template' accession name or on the DNA aligned sequence will display the corresponding footprintDB DNA motif information.

In the same way, left-clicking on the TF accession name in 'Binding proteins' or 'Interface sequences' columns will show the full information page for the TF:

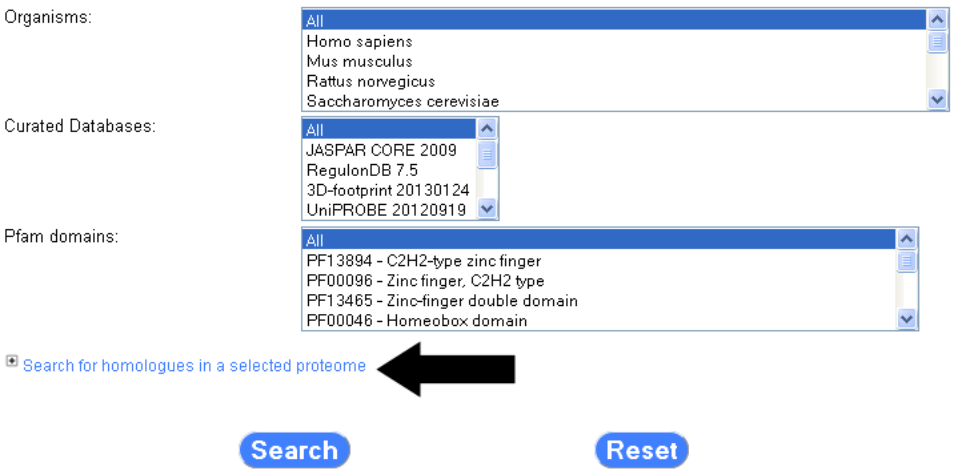



Transcription Factor
Accessions:
CAA74022 (JASPAR CORE 2009), bZIP910 (Athamap 20091028)
Names:
O22676_ANTMA, bZIP910
Organisms:
Antirrhinum majus
Libraries:
JASPAR CORE 2009 [1], Athamap 20091028 [2]
[1] Sandelin A., Alkema W, Engström P, Wasserman WW, Lenhard B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic acids research 32:D91-4 (2004). [Pubmed]
[2] Bülow L, Engelmann S, Schindler M, Hehl R. AthaMap, integrating transcriptional and post-transcriptional data. Nucleic acids research 37:D983-6 (2009). [Pubmed]
Length:
133
Pfam Domains:
13-61 Basic region leucine zipper
17-75 bZIP transcription factor
Sequence:
(in bold interface residues)
1 MASQQRSTSPGIDDDERKRKRKLSNRESARRSRMRKQQRLDELIAQESQMQEDNKKLRDT 60
61 INGATQLYLNFASDNNVLRAQLAELTDRLHSLNSVLQIASEVSGLVLDIPDIPDALLEPW 120
121 QLPCPIQADIFQC
Interface Residues:
21, 25, 26, 28, 29, 32, 33
Binding Motifs:
bZIP910(1) GATGACGTGGCm
bZIP910(2) GgrTGCTGACGT
MA0096.1 mTGACGT
Publications:
Martinez-Garcia J. F., Moyano E., Alcocer M. J. C., Martin C. Two bZIP proteins from Antirrhinum flowers preferentially bind a hybrid C-box/G-box motif and help to define a new sub-family of bZIP transcription factors.. Plant J. 13:489-505 (1998). [Pubmed]
Download

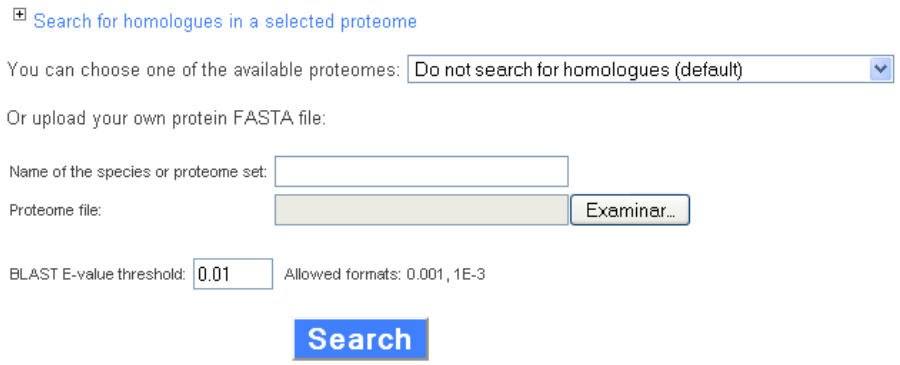
Other data shown are: the source database, organisms, Pfam domains, the set of interface residues -which are the key residues mediating specific DNA recognition-, STAMP E-value and DNA motif similarity score (sum of the Pearson correlation coefficients of the aligned DNA motif positions).

b. Find in a selected proteome homologous transcription factors that bind DNA motifs similar to the query

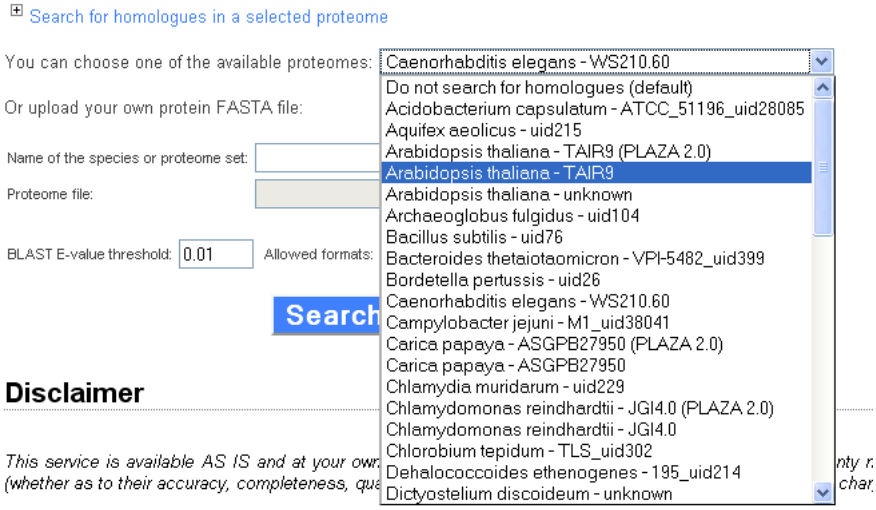
Please follow the steps explained in the former section 'Find transcription factors that bind DNA motifs similar to the query' until you see the search formulary. The menu **'Search for homologues in a selected proteome'** will be available at the bottom of the page:



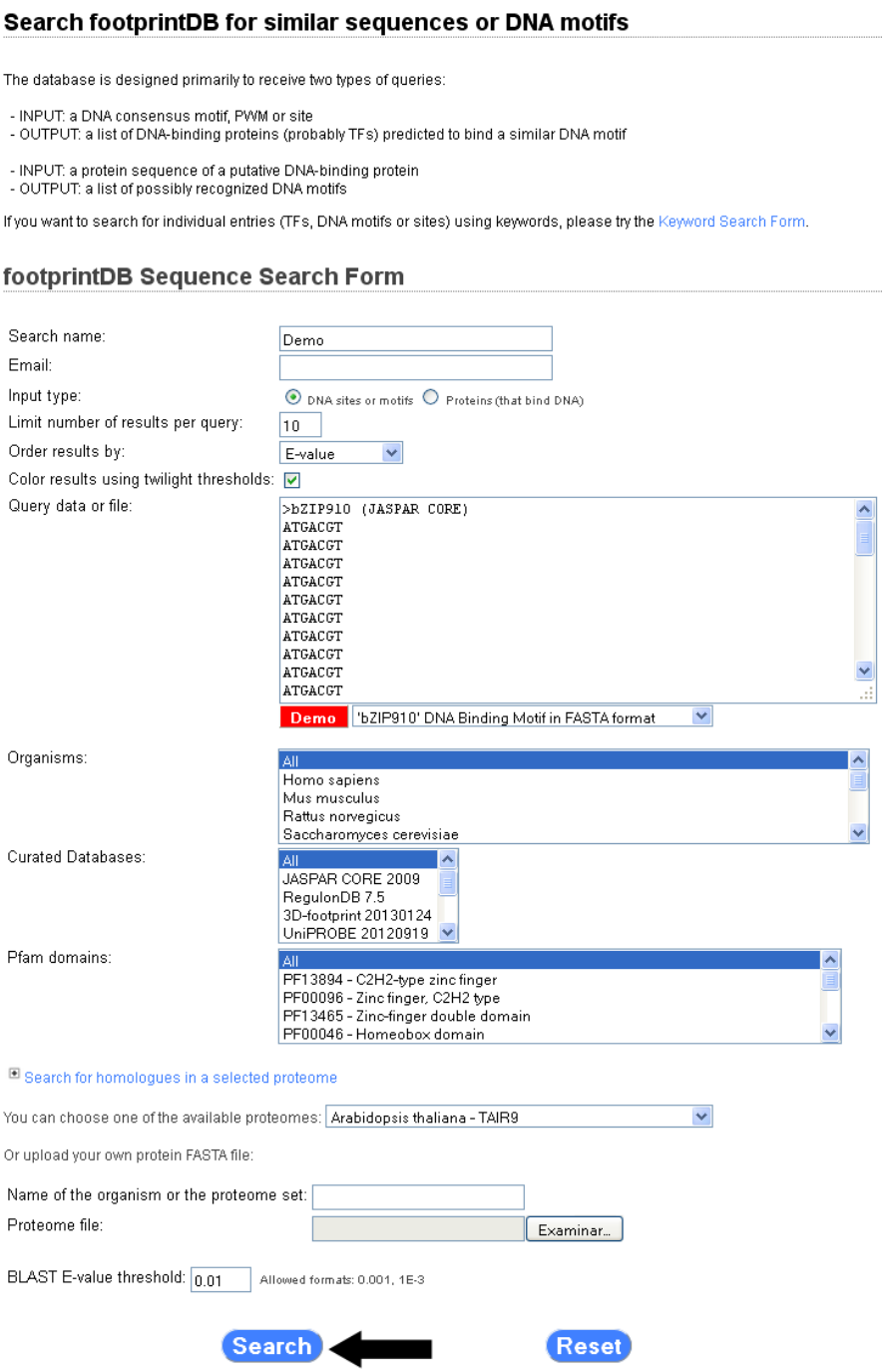
Click on the title **'Search for homologues in a selected proteome'** to expand the homologue search options:


Search for homologues in a selected proteome
You can choose one of the available proteomes:
Do not search for homologues (default)
Or upload your own protein FASTA file:
Name of the species or proteome set:
Proteome file:
Examinar...
BLAST E-value threshold:
0.01
Allowed formats: 0.001, 1E-3
Search

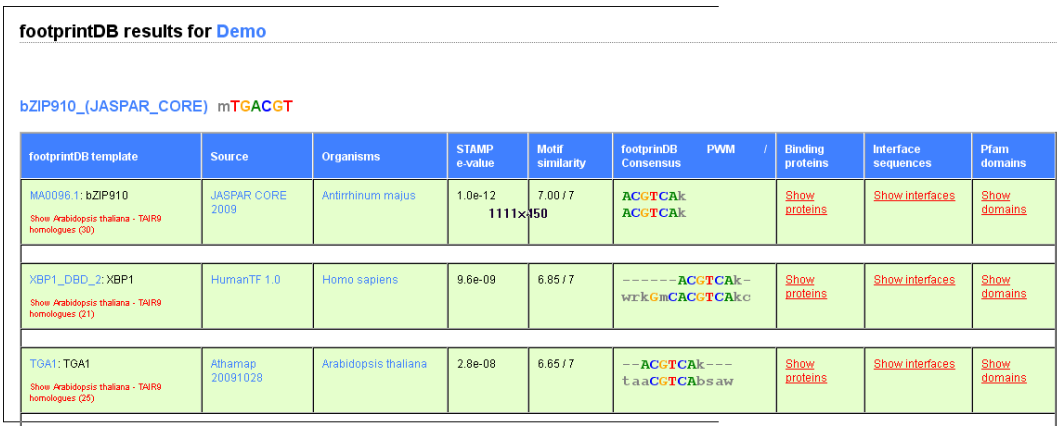
Now you might select a species to search for homologues in its proteome or either upload a file with a proteome file in FASTA format and choose a BLAST E-value threshold for the Blastp search against the proteome (Default 0.01).


Search for homologues in a selected proteome
You can choose one of the available proteomes:
Caenorhabditis elegans - WS210.60
Or upload your own protein FASTA file:
Name of the species or proteome set:
Proteome file:
BLAST E-value threshold:
0.01
Allowed formats:
Search
Disclaimer
Do not search for homologues (default)
Acidobacterium capsulatum - ATCC_51196_uid28085
Aquifex aeolicus - uid215
Arabidopsis thaliana - TAIR9 (PLAZA 2.0)
Arabidopsis thaliana - TAIR9
Arabidopsis thaliana - unknown
Archaeoglobus fulgidus - uid104
Bacillus subtilis - uid76
Bacteroides thetaiotaomicron - VPI-5482_uid399
Bordetella pertussis - uid26
Caenorhabditis elegans - WS210.60
Campylobacter jejuni - M1_uid38041
Carica papaya - ASGPB27950 (PLAZA 2.0)
Carica papaya - ASGPB27950
Chlamydia muridarum - uid229
Chlamydomonas reindhardtii - JGI4.0 (PLAZA 2.0)
Chlamydomonas reindhardtii - JGI4.0
Chlorobium tepidum - TLS_uid302
Dehalococcoides ethenogenes - 195_uid214
Dictyostelium discoideum - unknown

To start to search push **'Search'** button:



Search parameters are the same that in the previous example, but in this case we choose to include among the results the subset of *Arabidopsis thaliana* proteins which are presumably homologous to each of the reported DNA-binding proteins. Indeed we obtain the same previous results but in a slightly different order, with proteins with a higher number of homologues shown first (only the first 3 are discussed for brevity):



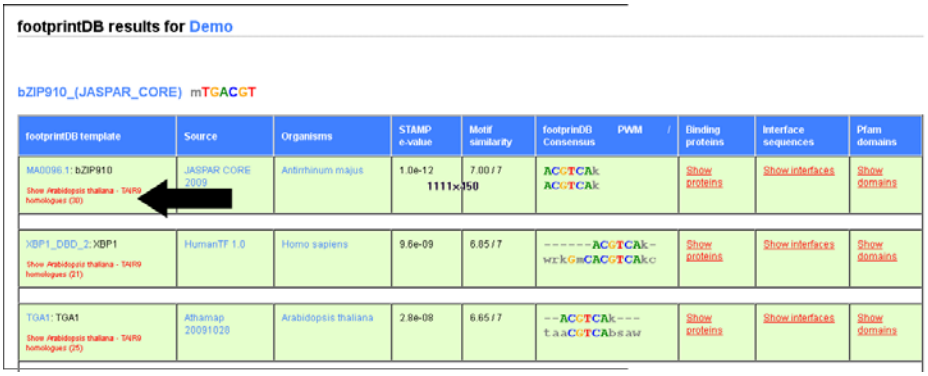
footprintDB results for Demo

bZIP910_(JASPAR_CORE) mTGACGT

footprintDB template	Source	Organisms	STAMP e-value	Motif similarity	footprintDB Consensus PWM /	Binding proteins	Interface sequences	Pfam domains
MA0096.1: bZIP910 Show Arabidopsis thaliana - TAIR9 homologues (30)	JASPAR CORE 2009	Antirrhinum majus	1.0e-12 1111x150	7.00 / 7	ACGTCAk ACGTCAk	Show proteins	Show interfaces	Show domains
XBP1_DBD_2: XBP1 Show Arabidopsis thaliana - TAIR9 homologues (21)	HumanTF 1.0	Homo sapiens	9.6e-09	6.85 / 7	------ACGTCAk- wrkGmCACGTCAkc	Show proteins	Show interfaces	Show domains
TGA1: TGA1 Show Arabidopsis thaliana - TAIR9 homologues (25)	Athamap 20091028	Arabidopsis thaliana	2.8e-08	6.65 / 7	--ACGTCAk--- taaCGTCAbsaw	Show proteins	Show interfaces	Show domains

Each row contains a TF that recognizes a motif similar to the query, and the motif alignment is also shown, as explained in the previous section. However,

the provided red link 'Show Arabidopsis thaliana – TAIR9 homologues' allows us to display a list of proteins, just beneath the name of each matched footprintDB name:



footprintDB results for Demo

bZIP910_(JASPAR_CORE) mTGACGT

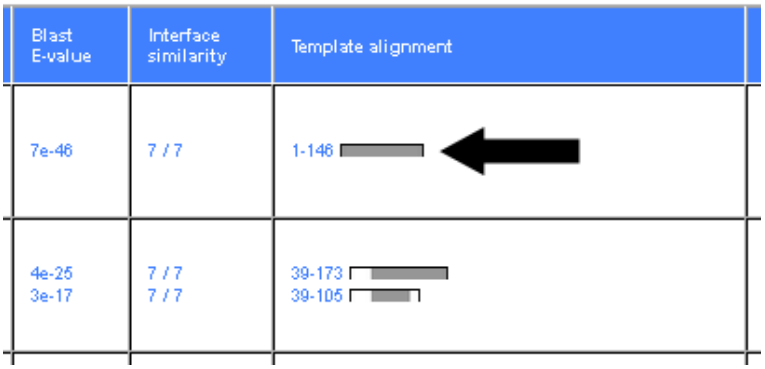
footprintDB template	Source	Organisms	STAMP e-value	Motif similarity	footprintDB Consensus PWM /	Binding proteins	Interface sequences	Pfam domains
MA0096.1: bZIP910 Show Arabidopsis thaliana - TAIR9 homologues (30)	JASPAR CORE 2009	Antirrhinum majus	1.0e-12 1111x150	7.00 / 7	ACGTCAk ACGTCAk	Show proteins	Show interfaces	Show domains
XBP1_DBD_2: XBP1 Show Arabidopsis thaliana - TAIR9 homologues (21)	HumanTF 1.0	Homo sapiens	9.6e-09	6.85 / 7	------ACGTCAk- wrkGmCACGTCAkc	Show proteins	Show interfaces	Show domains
TGA1: TGA1 Show Arabidopsis thaliana - TAIR9 homologues (25)	Athamap 20091028	Arabidopsis thaliana	2.8e-08	6.65 / 7	--ACGTCAk--- taaCGTCAbsaw	Show proteins	Show interfaces	Show domains

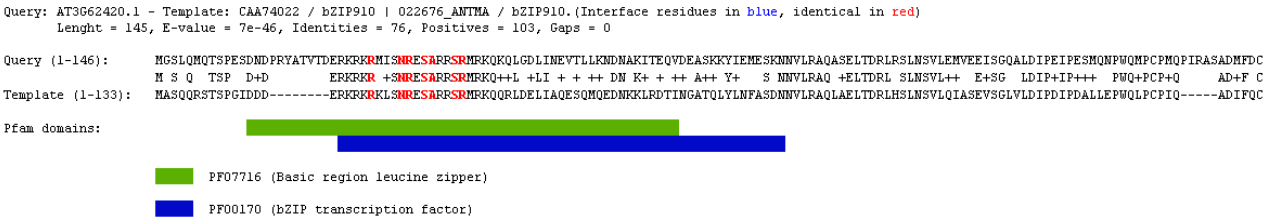
bZIP910_(JASPAR_CORE) mTGACGT

footprintDB template	Source	Organisms	STAMP e-value	Motif similarity	footprintDB Consensus PWM /	Binding proteins	Interface sequences	Pfam domains
MA0096.1: bZIP910 Show Arabidopsis thaliana - TAIR9 homologues (30)	JASPAR CORE 2009	Antirrhinum majus	1.0e-12	7.00 / 7	ACGTCAk ACGTCAk	Show proteins	Show interfaces	Show domains

Arabidopsis thaliana - TAIR9 protein / gene / interface	footprintDB template	Blast E-value	Interface similarity	Template alignment	Interface sequences	Related results
Protein: AT3G62420 (sequence) Interface: erkrkRmisNReSArrSRmrkqk	CAA74022 / bZIP910	7e-46	7 / 7	1-146	erkrkRklsNReSArrSRmrkqq	-
Protein: AT1G75390 (sequence) Interface: erkrkRkqsNReSArrSRmrkqk	CAA74022 / bZIP910 CAA74022 / bZIP910	4e-25 3e-17	7 / 7 7 / 7	39-173 39-105	erkrkRklsNReSArrSRmrkqq erkrkRklsNReSArrSRmrkqq	-
Protein: AT2G18160 (sequence) Interface: erkrkRmlsNReSArrSRmrkqk	CAA74022 / bZIP910	1e-23	7 / 7	29-123	erkrkRklsNReSArrSRmrkqq	-
Protein: AT4G34590 (sequence) Interface: qrkrkRmlsNReSArrSRmkkqk	CAA74022 / bZIP910	1e-21	7 / 7	26-108	erkrkRklsNReSArrSRmrkqq	-
Protein: AT1G68880 (sequence) Interface: erkrrRkvsNReSArrSRmrkqr	CAA74022 / bZIP910	2e-12	7 / 7	38-111	erkrkRklsNReSArrSRmrkqq	bZIP910_(JASPAR_CORE) (XBP1_DBD_2, TGA1)

Homologous proteins will be shown under each TF. Each new row contains data of one protein; left-clicking on the protein name will open a new window with protein sequence in FASTA format. **Left-clicking on the 'Blast E-value', 'Interface similarity' or 'Template alignment' columns will show the Blast alignment with the corresponding footprintDB protein sequence with coloured protein domains (Pfam version 24.0) highlighting in red the identical interface residues and in blue the rest of the interface.** The last column 'Related results' shows other footprintDB TF results which are presumably homologous to the same *Arabidopsis thaliana* protein.




Query: AT3G62420.1 - Template: CAA74022 / bZIP910 | 022676_ANTMA / bZIP910.(Interface residues in blue, identical in red)
Lenght = 145, E-value = 7e-46, Identities = 76, Positives = 103, Gaps = 0
Query (1-146): MGSLQMQTSPESDNDPRYATVTDERKRKRMISNRESARRSRMRKQKQLGDLINEVTLLKNDNAKITEQVDEASKKYIEMESKNNVLRAQASELTDRLRSLNSVLEMVEEISGQALDIPEIPESMQNPWQMPCPMQPIRASADMFDC
M S Q TSP D+D ERKRKR +SNRESARRSRMRKQ++L +LI + + ++ DN K+ + ++ A++ Y+ S NNVLRAQ +ELTDRL SLNSVL++ E+SG LDIP+IP+++ PWQ+PCP+Q AD+F C
Template (1-133): MASQQRSTSPGIDDD--------ERKRKRKLSNRESARRSRMRKQQRLDELIAQESQMQEDNKKLRDTINGATQLYLNFASDNNVLRAQLAELTDRLHSLNSVLQIASEVSGLVLDIPDIPDALLEPWQLPCPIQ-----ADIFQC
Pfam domains:
PF07716 (Basic region leucine zipper)
PF00170 (bZIP transcription factor)

3. Search protein sequences

a. Find transcription factors with similar sequences

Click on the ‘Search Sequences’ button in the Home page, click in the ‘Search’ option of the ‘Main menu’ or go directly to the url:
http://floresta.eead.csic.es/footprintdb/index.php?search

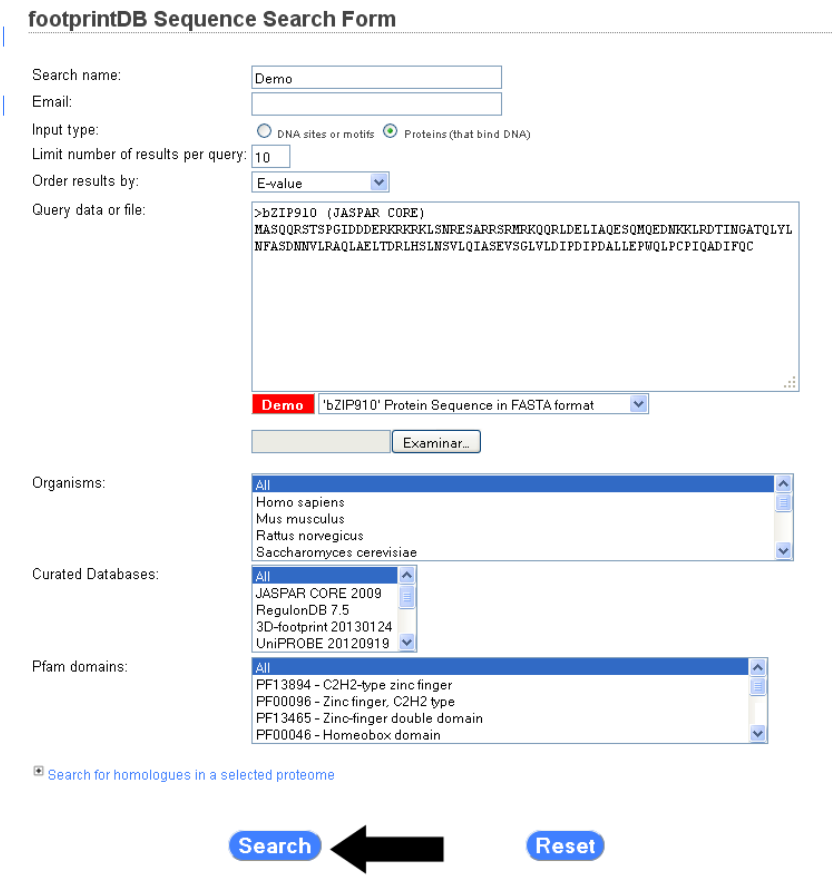
The search form fields and options are explained in the previous Section. In this case there is only a noticeable difference with respect to the input format of the sequence to search.

Query data or file: Enter your DNA sites or motifs in the text area or upload them from a file. The only valid format is FASTA . You can use sample data pushing the ‘Demo’ button.

Example of FASTA format.

```
>bZsP910 (JASPAR CORE)
MASQQRSTSPGIDDDERKRKRKLSNRESARRSRMRKQQRLDELIAQESQMQEDNKKL
RDTINGATQLYLNFASDNNVLRAQLAELTDRLHSLNSVLQIASEVSGLVLDIPDIPDALLEP
WQLPCPIQADIFQC
```

To start the search please click the **‘Demo’** button and then the **‘Search’** button:



In this search we query a protein sequence in FASTA format. In particular, we wish to search no more than 10 TFs with similar sequence and their associated DNA binding motifs without filtering organisms nor domains in all available databases.

We obtain the following results (only the first 3 are shown):

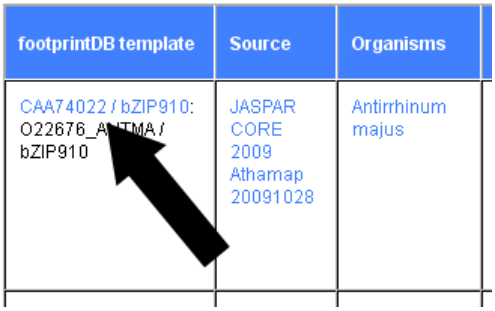
Query: bZIP910 (JASPAR CORE)

Sequence:
MASQQRSTSPGIDDDERKRKRKLSNRESARRSRMRKQQRLDELIAQESQMQEDNKKLRDTINGATQLYLNFASDNNVLRAQLAELTDRLHSLNSVLQIASEVSGLVLDIPDIPDALLEPWQLPCPIQADIFQC

footprintDB template	Source	Organisms	Blast e-value	Interface similarity	Interface residues	Interface sequence	footprintDB PWM / Consensus	Pfam domains
CAA74022 / bZIP910: O22676_ANTMA / bZIP910	JASPAR CORE 2009 Athamap 20091028	Antirrhinum majus	2e-94	7 / 7	21, 25, 26, 28, 29, 32, 33	...(16)erkrkRklsNReSArrSRmrk qq(38)...	bZIP910(1): GATGACGTGGCm bZIP910(2): GgrTGCTGACGT MA0096.1: mTGACGT	*13-61: PF07716 (Basic region leucine zipper)
CAA74023 / bZIP911: O22677_ANTMA / bZIP911	JASPAR CORE 2009 Athamap 20091028	Antirrhinum majus	2e-64	8 / 8	23, 26, 27, 28, 30, 31, 34, 35	...(18)erkrkRkqSNReSArrSRmrk qq(40)...	bZIP911(1): GrTGACGTGGCC bZIP911(2): GrTGACGTGTAC MA0097.1: GrTGACGTGkmC	*17-67: PF07716 (Basic region leucine zipper)
P25032: DNA-binding protein EMBP-1 / EMBP1_WHEAT / HBP-1a(1 / Histone promoter-binding protein 1a(1	JASPAR CORE 2009	Triticum aestivum	6e-08	10 / 10	253, 256, 257, 260, 261, 263, 264, 265, 267, 268	...(248)erelkRerRKqsNReSARrSR lrkqq(273)...	: GMCACGTGdv : wwcACGTGGW MA0128.1: WsACGTGG	*248-311: PF00170 (bZIP transcription factor)

We notice that the first result is the query itself (query is from JASPAR collection and is present in footprintDB) and the other results are transcription factors with similar interface sequences (results are ordered by Blastp E-value) and they have annotated also similar DNA motifs.

Each row contains a TF with similar sequence to the query. **Predicted DNA binding residues are shown coloured in the interface sequence and all the DNA motifs annotated for that TF are shown**. Left-clicking on the Blast E-value or the interface similarity score will show the alignment of the footprintDB TF sequence with the query. Left-clicking on the 'footprintDB template' TF accession name will display the full information about the TF:



Transcription Factor

Accessions: CAA74022 (JASPAR CORE 2009), bZIP910 (Athamap 20091028)

Names: O22676_ANTMA, bZIP910

Organisms: Antirrhinum majus

Libraries: JASPAR CORE 2009 [1], Athamap 20091028 [2]

[1] Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic acids research 32:D91-4 (2004). [Pubmed]

[2] Bülow L, Engelmann S, Schindler M, Hehl R. AthaMap, integrating transcriptional and post-transcriptional data. Nucleic acids research 37:D983-6 (2009). [Pubmed]

Length: 133

Pfam Domains:
13-61 Basic region leucine zipper
17-75 bZIP transcription factor

Sequence: (in bold interface residues)

```
  1 MASQQRSTSPGIDDDERKRKRKLSNRESARRSRMRKQQRLDELIAQESQMQEDNKKLRDT   60
 61 INGATQLYLNFASDNNVLRAQLAELTDRLHSLNSVLQIASEVSGLVLDIPDIPDALLEPW  120
121 QLPCPIQADIFQC
```

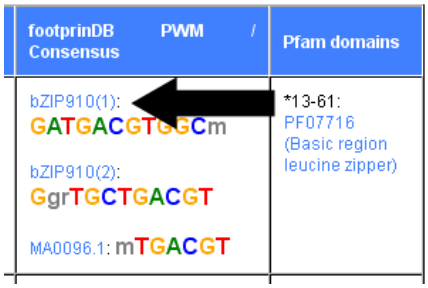
Interface Residues: 21, 25, 26, 28, 29, 32, 33

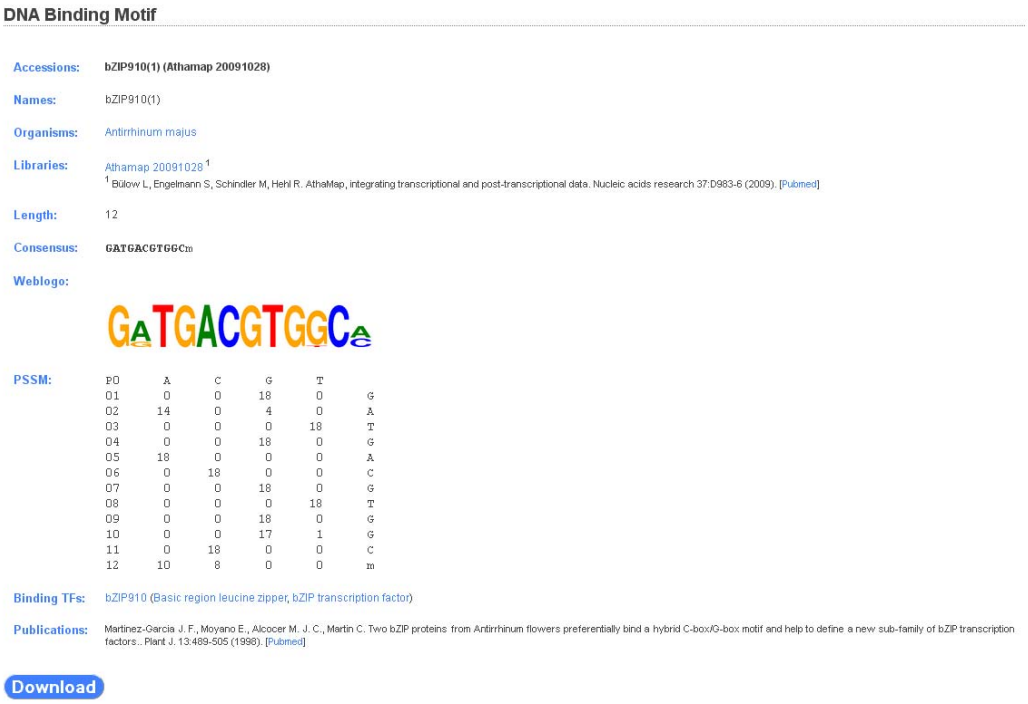
Binding Motifs:
bZIP910(1) **GATGACGTGGC**m
bZIP910(2) **G**g**rTGCTGACGT**
MA0096.1 m**TGACGT**

Publications: Martinez-Garcia J. F., Moyano E., Alcocer M. J. C., Martin C. Two bZIP proteins from Antirrhinum flowers preferentially bind a hybrid C-box/G-box motif and help to define a new sub-family of bZIP transcription factors.. Plant J. 13:489-505 (1998). [Pubmed]

Download

In the same way, left-clicking on DNA binding motif 'footprintDB PWM / Consensus' accession name will show the DNA binding motif information:



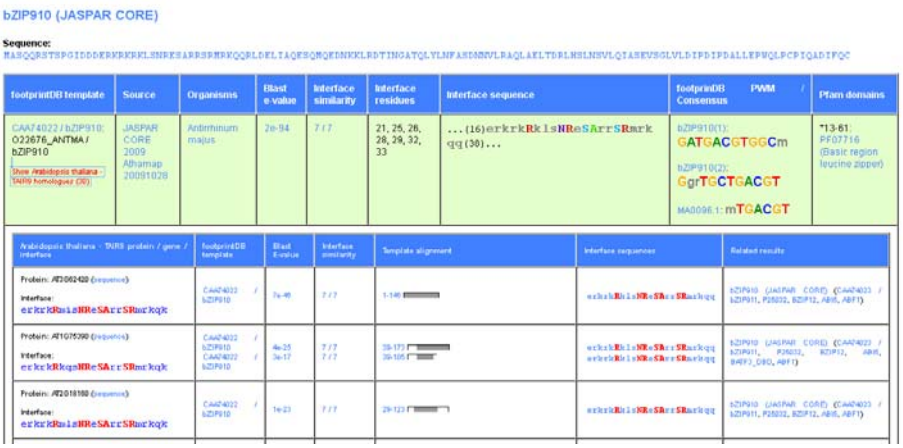

DNA Binding Motif
Accessions: bZIP910(1) (Athamap 20091028)
Names: bZIP910(1)
Organisms: Antirrhinum majus
Libraries: Athamap 20091028 [1]
[1] Bülow L, Engelmann S, Schindler M, Hehl R. AthaMap, integrating transcriptional and post-transcriptional data. Nucleic acids research 37:D983-6 (2009). [Pubmed]
Length: 12
Consensus: GATGACGTGGCm
Weblogo:
PSSM:
PO A C G T
01 0 0 18 0 G
02 14 0 4 0 A
03 0 0 0 18 T
04 0 0 18 0 G
05 18 0 0 0 A
06 0 18 0 0 C
07 0 0 18 0 G
08 0 0 0 18 T
09 0 0 18 0 G
10 0 0 17 1 G
11 0 18 0 0 C
12 10 8 0 0 m
Binding TFs: bZIP910 (Basic region leucine zipper, bZIP transcription factor)
Publications: Martinez-Garcia J. F., Moyano E., Alcocer M. J. C., Martin C. Two bZIP proteins from Antirrhinum flowers preferentially bind a hybrid C-box/G-box motif and help to define a new sub-family of bZIP transcription factors.. Plant J. 13:489-505 (1998). [Pubmed]
Download

Other data shown are: the source organism(s), Pfam domains, the set of interface residues -which are the key residues mediating specific DNA recognition-, Blastp E-value and interface similarity score.

b. Find in a selected proteome homologous transcription factors that bind TF sequences similar to the query

Please follow the steps explained in the former section 'Find transcription factors with similar sequences' until you see the search formulary. The menu **'Search for homologues in a selected proteome'** will be available at the bottom of the page. Then follow the same procedure explained in the previous section.

Homologous protein sequences from the selected genome will be shown and they can be accessed as previously explained.

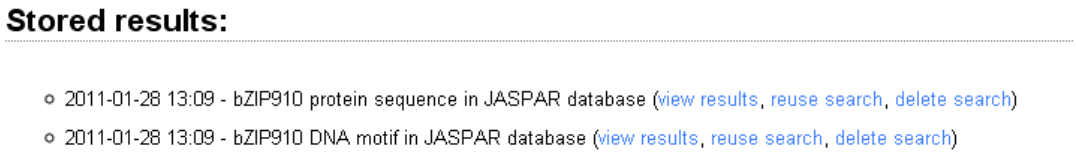


4. Retrieve stored searches

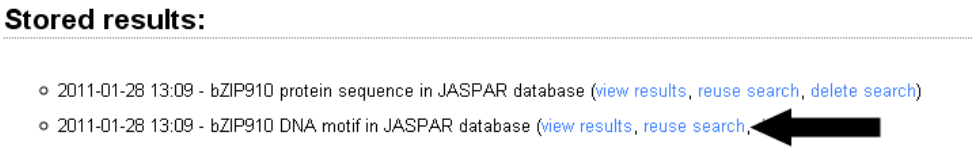
Registered users can access to a list of stored searches.

Log in and click on the 'Stored results' link in the 'User menu' on the left side:

A list of the performed searches will be shown:


Stored results:
2011-01-28 13:09 - bZIP910 protein sequence in JASPAR database (view results, reuse search, delete search)
2011-01-28 13:09 - bZIP910 DNA motif in JASPAR database (view results, reuse search, delete search)

Recent search results can be accessed by clicking on the 'view results' link. Old searches are deleted from the server; if you want to repeat one of these searches, click on the 'reuse search' link and the search formulary will be filled with the data of the old search:


Stored results:
2011-01-28 13:09 - bZIP910 protein sequence in JASPAR database (view results, reuse search, delete search)
2011-01-28 13:09 - bZIP910 DNA motif in JASPAR database (view results, reuse search,

5. Searching through the Web services interface

The footprintDB server can be accessed programmatically using a SOAP Web services interface. The following Perl source code illustrates how to make protein sequence, DNA motif and keyword queries:

```
#!/usr/bin/perl -w
use strict;
use SOAP::Lite;

my $footprintDBusername = ''; # type your username if registered
my ($result,$sequence,$sequence_name,$datatype,$keyword) = ('','','','','');
my $server = SOAP::Lite
-> uri('footprintdb')
-> proxy('http://floresta.eead.csic.es/footprintdb/ws.cgi');

## sample protein sequence
$sequence_name = 'test';
$sequence = 'IYNLSRRFAQRGFSPREFRLTMTRGDIGNYLGLTVETISRLLGRFQKSGMLAVKGKYITIEN';

$result = $server->protein_query($sequence_name,$sequence,$footprintDBusername);
unless($result->fault()){
      print $result->result();
}else{
      print 'error: ' . join(', ',$result->faultcode(),$result->faultstring());
}

## sample regulatory motif sequence
#$sequence = 'TGTGANNN'; # possible format
#$sequence = "TGTGA\nTGTGG\nTGTAG"; # another format
# transfac format for position weight matrices
$sequence= <<EOM;
DE 1a0a_AB
01 1 93 0 2
02 0 96 0 0
03 58 33 3 2
04 8 78 6 4
05 8 5 75 8
06 1 2 47 46
XX
EOM

$result = $server-
>DNA_motif_query($sequence_name,$sequence,$footprintDBusername);
unless($result->fault()){
      print $result->result();
}else{
      print 'error: ' . join(', ',$result->faultcode(),$result->faultstring());
}

## sample text query
$keyword = "myb";
$datatype = "tf"; # three alternative search types: “tf”,”motif”,”sites”
$result = $server->text_query($keyword,$datatype,$footprintDBusername);
unless($result->fault()){
      print $result->result();
}else{
      print 'error: ' . join(', ',$result->faultcode(),$result->faultstring());
}
```

Such queries generate XML output, that can also be programmatically parsed:

```
<?xml version="1.0"?>
<footprintdb>
<username></username>
<input protein name>test</input protein name>
<input protein
sequence>IYNLSRRFAQRGFSPREFRLTMTRGDIGNYLGLTVETISRLLGRFQKSGMLAVKGKYITIEN</input
protein sequence>
<results_summary>
footprintDB template    template common names   Source      Blast e-value
      Interface identity      Interface similarity    footprinDB Consensus
      Query alignment   Template alignment      Pfam domains
5989  FNR   RegulonDB 7.5     3e-39 7 / 7 7 / 7 tTGaTywayATCAA    1-62  174-235

...
Best result for the 'Q_evalue' classifier:     '5989' in position 1
Best result for the 'I_simil' classifier:'8472' in position 18
</results_summary>
<protein_sequences_fasta_format>
> 5989 | ECK120004795 | FNR
MIPEKRIIRRIQSGGCAIHCQDCSISQLCIPFTLNEHELDQLDNIIERKKP...
...
</protein_sequences_fasta_format>
<DNA_motifs_transfac_format>
DE  FNR | FNR
01     17      8     10     48      t
02      5      8      6     64      T
03     13     10     58      2      G
04     47     14      8     14      a
05     10      3     13     57      T
06     17     28      3     35      y
07     28     10     16     29      w
08     48     12     11     12      a
09     20     23     11     29      y
10     66      0     11      6      A
11      4      0      0     79      T
12      0     72      5      6      C
13     80      1      0      2      A
14     73      8      0      2      A
XX
...
DE  2isz_B | IDER_MYCTU / Iron-dependent repressor ideR
01      6      6      9     75      T
02      9      6      6     75      T
03     75      6      6      9      A
04      6      9     75      6      G
05      0      0     96      0      G
06      0      6     90      0      GXX
</DNA_motifs_transfac_format>
<citations>
footprintDB (http://floresta.eead.csic.es/footprintdb, PubMed=unpublished)

Please cite additional datasources as applicable:
JASPAR CORE (http://jaspar.cgb.ki.se/, PubMed=14681366*)
RegulonDB (http://regulondb.ccg.unam.mx, PubMed=23203884*,21051347)
3D-footprint (http://floresta.eead.csic.es/3dfootprint/, PubMed=19767616*)
UniPROBE (http://the_brain.bwh.harvard.edu/uniprobe, PubMed=21037262*,18842628)
DrosophilaTF (http://www.bioinf.manchester.ac.uk/bergman/data/motifs/,
PubMed=17238282*)
Athamap (http://www.athamap.de, PubMed=18842622*)
DBTBS (http://dbtbs.hgc.jp/, PubMed=17962296*)
```

HumanTF (http://www.cell.com/abstract/S0092-8674%2812%2901496-1,
PubMed=23332764*)
</citations>
<footprintdb>

<?xml version="1.0"?>
<footprintdb>
<username></username>
<input DNA motif name>test</input DNA motif name>
<input DNA motif sequence>
DE 1a0a_AB
01 1   93 0   2
02 0   96 0   0
03 58 33 3   2
04 8   78 6   4
05 8   5   75 8
06 1   2   47 46
07 1   2   84 9
XX
</input DNA motif sequence>
<results_summary>
footprintDB template   template common names   Source   STAMP e-value
      Motif similarity footprinDB Consensus   Interface residues   Pfam
domains
5957 PROTEIN (PHOSPHATE SYSTEM POSITIVE REGULATORY PROTEIN PHO4)   3D-
footprint 20130124   1.0e-12   7.00 / 7   CCmCGkG
...
Best result for the 'Q_evalue' classifier: '5957' in position 1
Best result for the 'I_simil' classifier:'5957' in position 1
</results_summary>
<protein_sequences_fasta_format>
> 8085 | 1a0a_A / 1a0a_B | PHO4_YEAST / PHOSPHATE SYSTEM POSITIVE REGULATORY
PROTEIN PHO4 / PHO4_YEAST / PHOSPHATE SYSTEM POSITIVE REGULATORY PROTEIN PHO4
MKRESHKHAEQARRNRLAVALHELASLIPAEWKQQNVSAAPSKATTVEAACRYIRHLQQNGST
...
</protein_sequences_fasta_format>
<DNA_motifs_transfac_format>
DE 1a0a_AB | PROTEIN (PHOSPHATE SYSTEM POSITIVE REGULATORY PROTEIN PHO4)
01   1   93   0   2   C
02   0   96   0   0   C
03   58   33   3   2   m
04   8   78   6   4   C
05   8   5   75   8   G
06   1   2   47   46   k
07   1   2   84   9   G
XX
...
</DNA_motifs_transfac_format>
</footprintdb>

<?xml version="1.0"?>
<footprintdb>
<username></username>
<input keyword>myb</input keyword>
<results_summary>
1272|AtMYB84(Athamap 20091028)
2555|TaMYB80(Athamap 20091028)
2728|CAA61021(JASPAR CORE 2009)|GAMYB(Athamap 20091028)
2814|CCA1(ArabidopsisPBM 20140210)
…
</results_summary>
<protein_sequences_transfac_format>
AC 1272|AtMYB84(Athamap 20091028)

XX
FA Myb-84
XX
SY AtMYB84; At3g49690.
XX
OS Arabidopsis thaliana
XX
SQ MGRAPCCDKANVKKGPWSPEEDAKLKSYIENSGTGGNWIALPQKIGLKRCGKSCRLRWLN
SQ YLRPNIKHGGFSEEEENIICSLYLTIGSRWSIIAAQLPGRTDNDIKNYWNTRLKKKLINK
SQ QRKELQEACMEQQEMMVMMKRQHQQQQIQTSFMMRQDQTMFTWPLHHHNVQVPALFRIKP
SQ TRFATKKMLSQCSSRTWSRSKIKNWRKQTSSSSRFNDNAFDHLSFSQLLLDPNHNHLGSG
SQ EGFSMNSILSANTNSPLLNTSNDNQWFGNFQAETVNLFSGASTSTSADQSTISWEDISSL
SQ VYSDSKQFF
XX
MX 687;
XX
RX PUBMED: 9628022
RL Romero I., Fuertes A., Benito M. J., Malpica J., Leyva A., Paz-Ares J. More
than 80R2R3-MYB regulatory genes in the genome of Arabidopsis thaliana. Plant J.
14:273-284 (1998).
XX
//
…
</protein_sequences_transfac_format>
<citations>
footprintDB (http://floresta.eead.csic.es/footprintdb, PubMed=24234003)

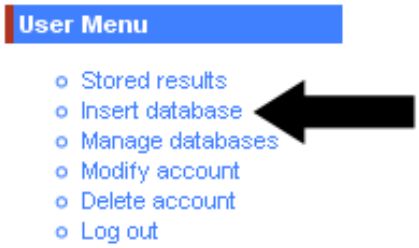
Please cite additional datasources as applicable:
JASPAR CORE (http://jaspar.genereg.net, PubMed=14681366*)
RegulonDB (http://regulondb.ccg.unam.mx, PubMed=23203884*,21051347)
3D-footprint (http://floresta.eead.csic.es/3dfootprint/, PubMed=19767616*)
UniPROBE (http://the_brain.bwh.harvard.edu/uniprobe, PubMed=21037262*,18842628)
DrosophilaTF (http://www.bioinf.manchester.ac.uk/bergman/data/motifs/,
PubMed=17238282*)
Athamap (http://www.athamap.de, PubMed=18842622*)
DBTBS (http://dbtbs.hgc.jp/, PubMed=17962296*)
HumanTF (http://www.cell.com/abstract/S0092-8674%2812%2901496-1,
PubMed=23332764*)
HOCOMOCO (http://autosome.ru/HOCOMOCO/, PubMed=23175603*)
ArabidopsisPBM (http://www.pnas.org/content/early/2014/01/29/1316278111,
PubMed=24477691*)
</citations>
</footprintdb>

Database insertion

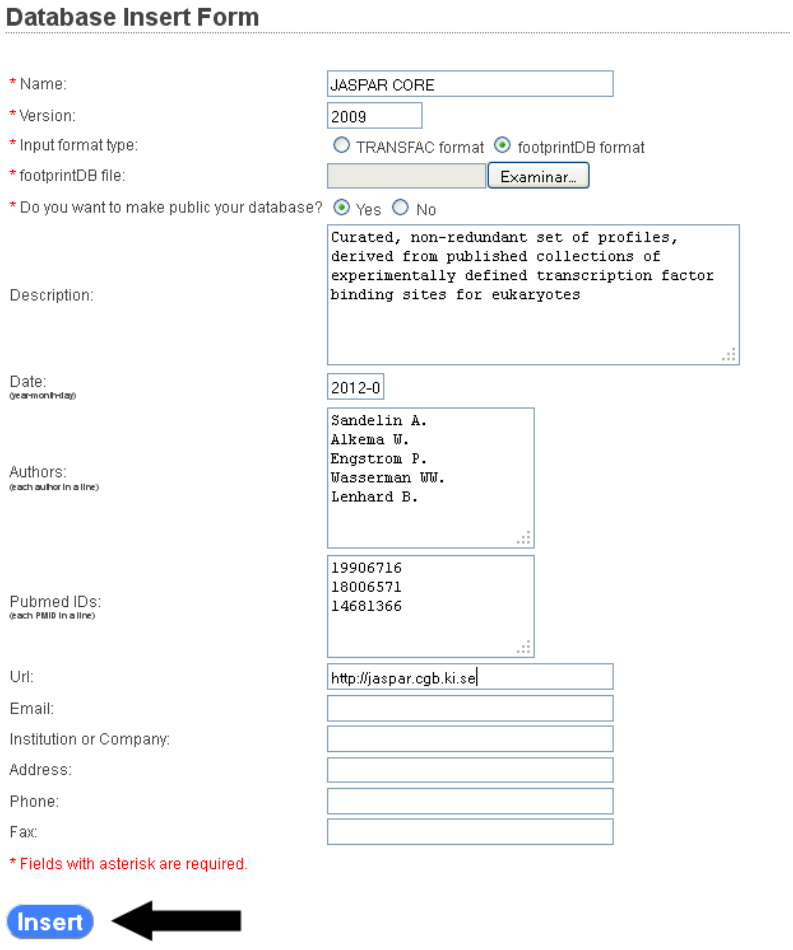
1. Insert a new database into footprintDB

If you are not registered, create a new account and log in.

Click on the option ‘Insert database’ in the ‘User Menu’ on the left or go directly to http://floresta.eead.csic.es/footprintdb/index.php?database_insert



Fill in all the fields about the new database and enter a file with the data in TRANSFAC of custom footprintDB format. These two formats will be explained in the next Section.



2. footprintDB and TRANSFAC data formats

First we will start explaining TRANSFAC format, the most used and standard format for DNA binding data and then we willl explain the unified footprintDB format that allows to store all the binding format in an unique file. The following format specifications must be followed to be able to insert data into footprintDB server.

a. TRANSFAC format

DNA binding data in TRANSFAC format is usually stored in three separated files: first one with TF sequences, second one with DNA motifs and matrices (PSSMs), third one with DNA single sites. The three files contain Identifiers and Accessions for each data entry, sequences or matrices and they have annotated the relationships among them. Besides other information like description, organism, annotations or literature references are usually included.

The three files have in common the following header:

```
VV  Header with library version
XX  End of field
//  End of entry
```

The DNA motif file has the following structure:

```
AC  Accession
XX
ID  Identifier
XX
NA  Main name
XX
DE  Description
XX
BF  Binding factor accession; Name; Species: ...
XX
P0  PSSM
01
...
XX
BS  Binding site data sequence; Accession;
XX
CC  Annotation
XX
RN [1] Reference number and Accession
RX  PUBMED: Pubmed ID
RA  Reference Authors
RT  Reference Title
RL  Reference Journal, Number, Issue, Pages (Year)
XX
//
```

The transcription factors file has the following structure:

```
AC  Accession
XX
ID  Identifier
XX
```

```
FA  Main name
XX
SY  Name synonyms (Separated by ';')
XX
OS  Organisms (Separated by ',')
XX
SQ  Sequence
XX
SC  Uniprot Uniprot ID
XX
FF  Annotation
XX
MX  Motif accession;
XX
BS  Binding site accession;
XX
RN  [1] Reference numer and Accession
RX  PUBMED: Pubmed ID
RA  Reference Authors
RT  Reference Title
RL  Reference Journal, Number, Issue, Pages (Year)
XX
//
```

The DNA sites file has the following structure:

```
AC  Accession
XX
ID  Identifier
XX
DE  Description
XX
OS  Organisms (Separated by ',')
XX
SQ  Sequence
XX
BF  Binding factor accession; Name; Species: ...
XX
MX  Motif accession;
XX
RN [1] Reference numer and Accession
RX  PUBMED: Pubmed ID
RA  Reference Authors
RT  Reference Title
RL  Reference Journal, Number, Issue, Pages (Year)
XX
//
```

b. footprintDB format

DNA binding data in footprinDB format is stored in a unique file containing TF sequences, DNA motifs and DNA sites information and relationships. Each entry is a single DNA motif that includes fields to annotate their DNA binding sites and related transcription factors.

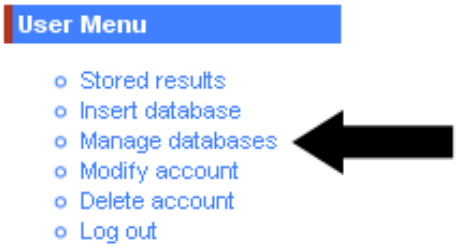
A footprintDB data file has the following fields and structure:

```
VV  Header with library data fields (Separated by ';')
VV  File: ; Name: ; Version: ; Date: ;
VV  Authors: ; Url: ; Email: ; Phone: ; Fax: ; Company: ; Address: ;
```

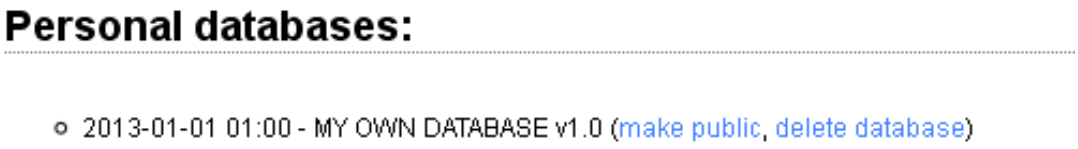
```
VV  Url: ; Pubmed: ; Description: ;
XX  End of section (header, motif, factor and site sections)
//  End of entry

# MOTIF SECTION:
MO  Accession
DE  Description
NA  Names (Separated by ';')
P0  PSSM
01
...
LN  Url
CC  Annotations (Separated by ';')
RX  PUBMED: Pubmed ID
RL  Reference details
XX

# FACTOR SECTION:
FA  Accession
DE  Description
NA  Names (Separated by ';')
SQ  Sequence
IN  (Blast prediction interface) Model: Residues; Total= ; Aligned= ;
IN  Identical= ; %ID= ; e-value= ; method=
SC  Uniprot Uniprot ID
OS  Organisms (Separated by ';')
LN  Url
CC  Annotations (Separated by ';')
RX  PUBMED: Pubmed ID
RL  Reference details
XX

# SITE SECTION:
SI  Accession
DE  Description
NA  Names (Separated by ';')
SQ  Sequence
LN  Url
CC  Annotations (Separated by ';')
RX  PUBMED: Pubmed ID
RL  Reference details
XX

# If the SITE has not Pubmed-Reference data, scripts will retrieve that
data from site's motif.

//
```

3. Manage own databases

Log in and click on the option ‘Manage databases’ in ‘User Menu’.



A list of all your previously inserted databases will be shown:



Two actions are available:

- **Make public**: if you select this option, your database will be public and open-access in footprintDB server.
- **Delete dabase**: you can remove previously inserted databases.

Bibliography

AlQuraishi, M. and H. H. McAdams (2011). "Direct inference of protein-DNA interactions using compressed sensing methods." Proc Natl Acad Sci U S A **108**(36): 14819-14824.

Bulow, L., S. Engelmann, et al. (2009). "AthaMap, integrating transcriptional and post-transcriptional data." Nucleic Acids Res **37**(Database issue): D983-986.

Contreras-Moreira, B. (2010). "3D-footprint: a database for the structural analysis of protein-DNA complexes." Nucleic Acids Res **38**(Database issue): D91-97.

Down, T. A., C. M. Bergman, et al. (2007). "Large-scale discovery of promoter motifs in Drosophila melanogaster." PLoS Comput Biol **3**(1): e7.

Jolma, A., J. Yan, et al. (2013). "DNA-Binding Specificities of Human Transcription Factors." Cell **152**(1): 327-339.

Lin, C. K. and C. Y. Chen (2013). "PiDNA: predicting protein-DNA interactions with structural models." Nucleic Acids Res.

Portales-Casamar, E., S. Thongjuea, et al. (2010). "JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles." Nucleic Acids Res **38**(Database issue): D105-110.

Robasky, K. and M. L. Bulyk (2011). "UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions." Nucleic Acids Res **39**(Database issue): D124-128.

Salgado, H., M. Peralta-Gil, et al. (2013). "RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more." Nucleic Acids Res **41**(Database issue): D203-213.

Sebastian, A. and B. Contreras-Moreira (2013). "The twilight zone of cis element alignments." Nucleic Acids Res **41**(3): 1438-1449.

Sierro, N., Y. Makita, et al. (2008). "DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information." Nucleic Acids Res **36**(Database issue): D93-96.